Universität St.Gallen

Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium

Bart Cockx, Michael Lechner, Joost Bollens

Januar 2020 Discussion Paper no. 2020-01

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor:	Vanessa Pischulti University of St.Gallen School of Economics and Political Science Department of Economics Müller-Friedberg-Strasse 6/8 CH-9000 St.Gallen Phone +41 71 224 23 07 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Müller-Friedberg-Strasse 6/8 CH-9000 St.Gallen Phone +41 71 224 23 07
Electronic Publication:	http://www.seps.unisg.ch

Priority to unemployed immigrants?
A causal machine learning evaluation of training in Belgium [1]

Bart Cockx[2], Michael Lechner[3], Joost Bollens

Author's address: Bart Cockx
Ghent University
Department of Economics
Sint-Pietersplein 6
BE-9000 Gent
Email bart.cockx@ugent.be
Website sites.google.com/site/bartcockxsite/

Author's address: Michael Lechner
Swiss Institute for Empirical Economic Research (SEW)
University of St.Gallen
Varnbüelstrasse 14
CH-9000 St.Gallen
Email michael.lechner@unisg.ch
Website michael-lechner.eu

Author's address: Joost Bollens
Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding (VDAB)
Keizerslaan 11
BE-1000 Brussels
Email joost.bollens@vdab.be

[1] We thank Michael Knaus and Anthony Strittmatter for carefully reading and commenting on a previous draft of the paper. The content of the paper was presented at the IZA workshop on the Evaluation of Labour Market Policies: New Data and New Approaches, in Copenhagen, September 2019, and at a research seminar of the economics department of the LMU University of Munich, December 2019. We thank participants, in particular Bas Van der Klaauw and Andreas Steinmayr, for helpful comments and suggestions. The usual disclaimer applies.

[2] Bart Cockx is also affiliated with IZA Bonn, CESIfo, Munich, IRES, Université catholique de Louvain and ROA, Maastricht University.

[3] Michael Lechner is also affiliated with CEPR, London, CESIfo, Munich, IAB, Nuremberg, IZA, Bonn, and RWI, Essen. Michael Lechner gratefully acknowledges financial support from a grant of the NRP 75 of the Swiss Science Foundation (grant SNSF 407540_166999).

Abstract

We investigate heterogenous employment effects of Flemish training programmes. Based on administrative individual data, we analyse programme effects at various aggregation levels using Modified Causal Forests (MCF), a causal machine learning estimator for multiple programmes. While all programmes have positive effects after the lock-in period, we find substantial heterogeneity across programmes and types of unemployed. Simulations show that assigning unemployed to programmes that maximise individual gains as identified in our estimation can considerably improve effectiveness. Simplified rules, such as one giving priority to unemployed with low employability, mostly recent migrants, lead to about half of the gains obtained by more sophisticated rules.

Keywords

Policy evaluation, active labour market policy, causal machine learning, modified causal forest, conditional average treatment effects

JEL Classification

J68

1 Introduction

Unemployment remains an important economic and social concern in Europe, even though in the European Union (EU) the (overall) unemployment rate has steadily decreased from 10.9% in 2013 to 6.8% in 2018 (Eurostat). This is particularly true for some vulnerable groups, such as youth, older workers, and migrants. Policy makers therefore have continued interest in getting a better understanding of which labour market policies work for which type of unemployed. Such understanding helps to improve the counselling process, the design and the allocation of active labour market policies.

However, uncovering heterogeneity in the effectiveness of policies is challenging from an econometric point of view, because it requires estimators that are at the same time sufficiently flexible and sufficiently precise when predicting causal effects at such a fine grained-level. Recent developments in causal machine learning (CML) addressed this problem and offered promising solutions. In this paper we use such a CML approach for multiple treatments to evaluate the heterogeneity in the effectiveness of training programmes in Flanders, a region in the North of Belgium. We then use our estimates to uncover specific heterogeneity and to show the extent the public employment service (PES) can enhance the effectiveness of these programmes by changing the assignment of unemployed job seekers to these programmes.

Machine learning methods are traditionally used for prediction (e.g. Hastie, Tibshirani, and Friedman, 2009). More recently, these methods have been modified such that are useful for causal inference as well (see Athey 2019, and Athey and Imbens, 2019, for overviews). This literature shows how the counterfactual causal problem (e.g. Imbens and Wooldridge 2009) can be transformed into a combination of specific prediction problems. For this paper, these methods are of interest because they provide a way to systematically uncover the underlying heterogeneity of the causal effects, a goal for which traditional econometric methods fail to provide a systematic solution.

In this paper identification of the causal effect relies on the assumption of unconfoundedness. Knaus, Lechner and Strittmatter (2018) evaluate the performance of various CML methods for binary treatments suitable under unconfoundedness using an Empirical Monte Carlo approach (see e.g. Huber et al. 2013; Lechner and Wunsch 2013). As opposed to a standard Monte Carlo analysis, such an approach informs the data generating process (DGP) as much as possible by real data and reduces the synthetic components in the DGP to a minimum. The real data are taken from the Swiss social security records that were used to evaluate a job search programme of the unemployed (Knaus, Lechner and Strittmatter 2017). Knaus, Lechner and Strittmatter (2018) conclude that the Forest based ML methods, in particular the Generalized Forest by Athey, Tibshirani, and Wager (2018), belong to the best performing estimators if explicitly adjusted to take account of confounding. Subsequently, Lechner (2018) proposes the Modified Causal Forest (MCF) estimator. It builds on the estimators proposed by Wager and Athey (2018) and Athey et al. (2018). One innovation is that Lechner (2018) improves the objective function used to build the trees of the Causal Forest. The second innovation proposed is to use weight-based inference methods as a computationally cheap and reliable device to estimate the precision of the estimated treatment effects at the various aggregation levels of interest, from the individualized to the (grouped) average treatment effects. Based on an Empirical Monte Carlo analysis, Lechner (2018) shows that the MCF estimator outperforms previously suggested estimators in nonexperimental settings. Since the MCF allows to effectively address programme heterogeneity as well as individual heterogeneity at various levels, has attractive theoretical properties and seems to perform well in practice, it is our estimator of choice.

To the best of our knowledge, this paper is one of the first papers that applies CML methods to analyse treatment heterogeneity in the evaluation of active labour market policies. It appears also to be the first paper to use perform such analysis in a multiple treatment context. Knaus, Lechner and Strittmatter (2017) use Lasso based methods to evaluate the effect heterogeneity of a (single) job search programme in Switzerland using admin data from 2003. They find substantial effect heterogeneity, but only during the first six months after the start of programme participation. Bertrand, Crépon, Marguerie and Premand (2017) apply the Causal Random Forest method of Wager and Athey (2018) within a randomized controlled trial (RCT) to evaluate the programme heterogeneity of a temporary public works programme in a less developed country. Their analysis reveals important heterogeneity, but again mostly during programme participation. Faltings, Krumer, and Lechner (2019) use the MCF approach to analysis regional favouritism of referees in the top Swiss soccer league.

This paper uses administrative data of the Flemish PES. The analysis is based on the population of about 60.000 individuals aged between 21 and 55 who started claiming unemployment insurance benefits after an involuntary lay-off between December 2014 and June 2016. To follow-up labour market outcomes during at least 2.5 years (until September 2019), we evaluate the impact of participating in training programmes that were entered within the first nine months of the unemployment spell. We focus the analysis on three training programmes: short-term (less than 6 months) vocational training (SVT), longer-term (between 6 and 10 months) vocational training (LVT) and orientation training (OT) that aims at helping to determine a clear professional goal.[1]

The interest in using data from the Flemish PES is threefold. First, the type of training offered to the unemployed are like those offered in most other EU countries, so that findings for Flanders are of general interest. Second, the administrative data is very informative. In addition to rich socio-demographic information, the data contains extensive information on labour market histories (including sickness and past programme participation) of individuals since

[1] Other programs were not considered for various reasons: (i) they did not pass a placebo test; (ii) Dutch language training because of lack of comparison observations (common support problem); (iii) they lasted too long (more than 10 months) in view of the time horizon of 2.5 years; (iv) on-the-job training, because participants were selected after hiring; (v) they were too small and too heterogeneous to be aggregated in a meaningful group.

1991. This makes the assumption of unconfoundedness, which is used for identification, arguably plausible. In fact, in a placebo exercise unconfoundedness could not be rejected. Third, the Flemish PES displayed a high willingness to employ CML methods in the future programme evaluations and assignments.

The main findings of this paper can be summarized as follows. There is a clear dominance ordering in terms of the average effectiveness of the three programmes considered, both in the short-run (lock-in effects) as in the medium run (post-programme effect): SVT clearly performs best, followed by LVT. Although OT also shows positive post-programme effects, the lock-in effects are so large and enduring, that the overall effect in our observation period is negative. After 2.5 years, participation in SVT increases on average the time spent in employment by 3.4 months relative to no participation. For LVT this gain is only 1 month, while participation in OT decreases the number of months in employment by 1.4. There is considerable heterogeneity in the aforementioned effects. The effects are especially higher for recent migrants with a low proficiency in Dutch, which is the official language in Flanders. The level of education is also an important determinant of the programme effects. The effectiveness of these programmes decreases with the level of education and with the extent to which this education is oriented directly towards the labour market: The programmes work best for those who dropped out of the general track that prepares for university and work less well for those who drop out or complete a vocational or technical schooling track. Interestingly, in contrast to previous findings, heterogeneity in these dimensions is not only present during the lock-in phase, but also in the post-treatment phase. Nevertheless, overall heterogeneity is more prevalent in the lock-in phase and is present also in other dimensions such as age, unemployment duration and past unemployment experience. This follows from the fact that during the lock-in phase the implicit costs of participating in a programme, i.e. not searching for a job, are generally lower for those who are less likely to find a job anyway.[2]

Finally, we study to which extent the Flemish PES could improve the effectiveness of their training component of the ALMP by changing the allocation of programme participants according to the individualized effectiveness of these programmes. We consider two main scenarios: one in which we impose no capacity constraints, so that an unlimited number of unemployed could be allocated to programmes, and one in which the capacity of the programmes was fixed to the observed level. In the first case, changing the allocation could increase the time spent in employment by about 3.3 months. In the second scenario the overall gain is limited, i.e. only about 9 days for all, essentially because the capacity of the programmes is too small to have a large impact on overall employment. However, the small population of participants would gain substantially, namely about 5 months within a 30-month window. Using simpler allocation rules based on employability or past unemployment histories also leads to gain, but they reduce to about half of the gains of the optimal allocation.

The rest of the paper is structured as follows. In Sections 2 and 3 we describe the institutional setting and the data that are used in the analysis. Section 4 discusses the econometric methods. Section 5 presents the results with a focus on the analysis of effect heterogeneity. Section 6 simulates several alternative assignment rules. It is followed by some robustness analysis and concluding remarks.

2 The institutional setting

Belgium is a federal state in which many competences have been decentralized. Generally, location-based matters, such as employment policies, are decentralized to the three regions

[2] This fact also explains the result in Knaus et al. (2017) who find larger effects of job search programs for foreigners, but only during the lock-in phase.

(Flanders, Wallonia, and Brussels) and language-based matters, such as education, to the three communities (the Flemish, French and the small German one). National defence, justice and Social Security are typical competences that remained at the federal level. The rules and payment of unemployment insurance (UI) are thus determined at the federal level. The Regional Public Employment Services (PES) oversee job search assistance, intermediation services and the provision of active labour market policies for the unemployed. While the monitoring and the associated sanctioning of labour market availability and job search effort were until the end of 2015 executed by the federal unemployment agency (RVA/ONEM), based on information transmission by the regional PES, these tasks have been transferred to the regional PES since 2016 (in Brussels since 2017).

In general, in Belgium an unemployed worker who is seeking employment is entitled to non-means tested unemployment benefits (UB) in two cases. First, graduates from high school or higher education who are younger than 25 can start claiming benefits one year after graduation if they have been continuously searching for a job, but did not find one in this period. Second, workers are eligible for UB after involuntary lay-off to the extent that they have sufficiently contributed to the UI. This is the scheme on which we focus in this paper. Unlike in other countries, the benefits are paid out without time limit. Laid-off workers the UB is related to prior earnings, but bracketed by a cap and a floor. The replacement rate is initially 65% (with a maximum of €1,736/month), but declines with unemployment duration. It reduces to 60% after 3 months, and then further after one year, depending on the status within the household and prior work experience. After 4 years all UBs attain the minimum which depends on the household status: €1,316/month for heads of household, €1,078/month for singles, and €561 for dependents, before taxes.[3]

[3] Amounts for July 2019 (https://www.rva.be/nl/documentatie/infoblad/t67).

The regional PES provides employment services to unemployed and checks their availability for the labour market, already to some extent before the 2016 reform, and fully afterwards. Workers younger than 55 should be both passively and actively available for the labour market. Being (passively) available means that these workers should register as job seeker, show up at meetings convoked by PES counsellors and at job interviews with employers, and accept "suitable" job offers, programme participation and counselling. Being actively available means that they should be seeking a job. In Flanders passive availability was monitored more intensively than active availability, both before and after the 2016 reform. However, both forms of monitoring can be characterised as relatively "loose" from an international perspective (Langenbucher, 2015).

An unemployed worker who registers at the regional PES in Flanders (VDAB) is invited to an intake meeting or phone call with a counsellor. At this intake meeting information is provided with respect to the rights and duties. In the period of analysis, the computer system of the PES informed the caseworker whether the unemployed job seeker should subsequently be invited for a new meeting. During these meetings the caseworker determines whether the unemployed worker is in need of services of the PES. Amongst them, there are variety of active labour market policies (ALMP), such as *orientation training* (OT) - helping job seekers in identifying the professions to aim at - *vocational training* (VT) - learning specific competences required in certain professions - *language training* (DLT) - Dutch for foreigners -, *on-the-job training* (OJT), or *intensive counselling* scheme (IC). In the next section we explain which programmes were retained for the analysis and why.

Internationally comparable statistics on the importance of ALMP are not available at the regional level, but expenditures in Belgium are not very different from a typical OECD country (see OECD.stat). Labour market services includes counselling and job search assistance. The combination of these services with training is therefore the best proxy in these statistics for the

ALMP provided by a PES. In 2016, the Belgian expenditures on these posts sum to 0.35% of GDP. This is somewhat higher than the OECD average of 0.25% and higher than the 0.30% for the neighbouring country, the Netherlands. The other neighbours, France, and Germany, spend more, respectively 0.54% and 0.55% of GDP.

3 Data

3.1 Population of interest

The data for the analysis is drawn from administrative files on all individuals who registered between January 1991 and February 2019 as unemployed job seeker at the Flemish PES. These files contain rich socio-demographic information, as well as individual employment and job search histories since 1991. From this database we select 148,942 individuals who started claiming UI after an involuntary lay-off between December 2014 and June 2016. We do not retain individuals who enter unemployment after June 2016 as to allow follow-up of participants for a sufficiently long period. Since we retained programmes that commenced up to 9 months after the start of the unemployment spell and since the observation period ends in September 2019, labour market outcomes can be observed for up to 30 months after the programme start.

We exclude school-leavers claiming UB (see Section 3) as well as individuals younger than 21. We also exclude workers with disabilities and those older than 55 at the start of the unemployment spell, because these groups may not need to be fully available for the labour market or may benefit from alternative policies. We also dropped individuals not living in Flanders and those who died during the period of analysis. 73,582 individuals are retained after imposing these selection criteria.

3.2 Programmes

Table 3.1 reports how this population is divided up into four subgroups: (1) 56,324 individuals who did not participate in any ALMP within the first nine months of unemployment, i.e. the not yet treated group (Sianesi, 2004); (2) 3,640 individuals who started within the first nine months an ALMP that is retained for the main analysis; (3) 13,618 individuals entered within the first nine months an ALMP not retained for the evaluation, because (i) the placebo tests suggested concerns about selection bias when evaluating intensive counselling, (ii) in Dutch language training almost only foreigners with limited language skills participated, so that too few comparison observations were available, (iii) on-the-job training is not comparable to the other ALMP, as it is assigned to individuals who have already found a job and (iv) other small ALMP could not be aggregated in meaningful groups sufficiently large for an empirical analysis, or they belonged to a category that lasted too long on average (10 months or more) for an evaluation of the medium run impacts within the 30 months observation period.

Programme participants are classified according to the first programme they participate in. The programmes considered in the analysis are the following. First, we distinguish between *short*- (less than 6 months, 3.8 months on average) and *long* (more than 6 and less than 10 months, 7.8 months on average) *vocational training* programmes (SVT and LVT). Very long vocational training programmes, lasting 10 months or more, were reclassified into *other ALMP,* and subsequently dropped from the analysis as to ensure a sufficiently long follow-up. Since no information on planned duration is available, the duration of vocational training (VT) is determined based on the average realized duration in the corresponding sector.

Third, *orientation training* (OT) aims at helping to determine a clear professional goal. This programme is relatively short. It lasts on average only about one month. However, within three months after the end 45% percent of the participants enter another ALMP, presumably to support the orientation that they have chosen.

Table 3.1: Importance of ALMP by type for entry cohorts in unemployment between December 2014 and June 2016

Type of assistance and ALMP	Average programme duration (months)	Number of individuals	Fraction
No ALMP participation within first 9 months (NOP)[1]	-	56,324	76.5%
ALMP within first 9 months retained in main analysis		3,640	4.9%
1. Short (< 6 months) vocational training (SVT)[2]	3.83	1,305	1.8%
2. Long (< 10 months) vocational training (LVT)[2]	7.18	1,220	1.7%
3. Orientation training (OT)	1.05	1,115	1.5%
ALMP within first 9 months excluded from analysis		13,618	18.5%
1.Intensive counselling (IC)	-[4]	3,695	5.0%
2. Dutch language training (DLT)	2.56	991	1.3%
3. On-the-job training (OJT)	-	2,045	2.8%
4. Other ALMP, including *very long* VT[2,3]	-	6,887	9.4%
Total		73,582	100.0%

Note: Individuals aged between 21 and 55 years at registration who started claiming an UI benefit after lay-off in the period December 2014-June 2016. The following individuals were excluded: (i) those not living in Flanders; (ii) those with some disability; (iii) those who died during the period of analysis;
[1] This group may enter programmes beyond the first 9 months.
[2] No information on planned duration available. Duration of vocational training (VT) is determined as average realized duration in the corresponding sector. Since the number of VT in some sectors was too small, some of them were aggregated. Eventually, 31 sectors are distinguished, all containing at least 19 individuals.
[3] This group contains various types of small, heterogeneous ALMP as well as 1,492 individuals who participated in vocational training lasting 10 months or more.
[4] The administrative files only record the administrative end of the contract as determined by the service provider to which the IC is contracted out. The duration of the service provision is not known, but must be less than the contract duration which lasts 8.82 months on average.

The dataset contains 45 ordered and 9 categorical conditioning variables (with 3 to 44 unordered categories). All time-varying variables are measured at the start of the unemployment spell. Taking into account that categorical variables are transformed into dummy variables in a regression-type setting,[4] this would correspond with 175 variables in a regression framework and many more if one aims at avoiding parametric restrictions by the inclusion of interaction and higher order terms for which the MCF will automatically account for.

These conditioning variables provide information about personal socio-demographic characteristics, labour market history, including sickness, within the preceding 2, 5 and 10 years, the ALMP participation history during previous unemployment spells, information about the job seeker's job preferences and the corresponding professional experience, the calendar

[4] This recoding is not needed for the MCF as it treats categorical variables directly, like Random Forests (as in Chou 1991; Hastie et al. 2009/2013, p. 310).

month in which unemployment was entered (19 indicators) and the day at which the ALMP started (or was predicted to start in case of no participation)[5] in the unemployment spell (maximum 274 days).

Table 3.2 reports summary statistics for a selected set of conditioning variables (panel A) and outcomes (panel B): the sample means by programme status and the standardized differences (in %) of for each programme status (SVT, LVT, OT) relative to the NOP group that did not participate in any ALMP within the first 9 months of the unemployment spell. A full description of all explanatory variables and the corresponding statistics can be found in Appendix A. The standardized differences are often larger than 20%, a number that Rosenbaum and Rubin (1985) consider 'large'. This signals that the conditioning variables of participants in ALMP are very unbalanced relative to the NOP group and that controlling for selection bias is crucial in this setting.

Table 3.2: Means and standardized differences for selected variables

Variable	No ALMP participation (NOP)	Short vocational training (SVT)		Long vocational training (LVT)		Orientation training (OT)	
A. Conditioning variables	Sample mean (*standardized difference*100 relative to NOP*)[1]						
Woman	0.49	0.31	*(36)*	0.40	*(16)*	0.46	*(3)*
Age (in years)	35	34	*(12)*	34	*(12)*	34	*(16)*
Proficiency in Dutch (0-3)[2]	2.4	2.5	*(8)*	2.7	*(40)*	2.6	*(27)*
Months unemployed in last 10 years	18	19	*(5)*	16	*(11)*	17	*(4)*
Months unemployed in last 2 years	3.9	3.8	*(2)*	3.0	*(18)*	3.2	*(14)*
Education level (1 to 13)	7.2	5.9	*(38)*	7.9	*(22)*	7.2	*(1)*
B. Outcomes							
# of months employed 10 months after start ALMP[3]	4.0	3.9	*(2)*	2.8	*(33)*	2.4	*(45)*
# of months employed 20 months after start ALMP[3]	9.8	11	*(16)*	9.4	*(5)*	7.8	*(27)*
# of months employed 30 months after start ALMP[3]	16	18	*(24)*	17	*(11)*	15	*(12)*
Number of observations	59964	1305		1220		1115	

Notes: [1] The standardized difference is defined as $\left|\bar{x}^j - \bar{x}^{NOP}\right| / \sqrt{[Var(x^j) + Var(x^{NOP})]/2} * 100$, where $\bar{x}^j$ and $Var(x^j)$ are the sample mean and variance of the variable x^j for $j \in \{SVT, LVT, DLT, IC\}$.
[2] Proficiency in Dutch = 0 if no knowledge; = 1 if limited; = 2 if good; =3 if very good.
[3] For non-participants in ALMP (NOP) the date at which the ALMP starts is predicted (See Section 4.4).

We observe that there are many more men than women participating in *vocational training*. Participants in vocational training are somewhat more proficient in Dutch, especially those

[5] More on this in Section 4.4.

in LVT. Participants in SVT have comparable unemployment experience as non-participants, while those in LVT have clearly been less unemployed both in the last 2 and 10 years. Participants in SVT are less educated than non-participants, while those in LVT are on average more educated.

Participants in *orientation training* are much more likely to have a good knowledge of Dutch. Moreover, they have been on average much less unemployed in the last two years. They are more likely to have a medium level of education. This profile seems to match therefore the profile of a medium skilled worker who has been employed in a routine job and has been displaced in the gradual tendency of more polarization of the labour market (see e.g. Autor et al. 2003; Goos et al. 2009). These workers typically require to be re-oriented to another profession, because they typically have skills that are no longer in demand.

Panel B of Table 3.2 reports the summary statistics for three main outcome variables, namely cumulative number of months that a worker is employed 10, 20 and 30 months after the start of the ALMP. In the empirical part, the last one and additional outcome variables will be considered. Since participation in ALMP can last up to 10 months on average (for long VT), the effects after 10 months measure *lock-in* effects for some programmes, while after 30 months the post-program effect, if present, adds to this lock-in effect.

It can be deduced from panel B of Table 3.2 that the outcomes vary substantially by programme status. However, in view of the important variability of the conditioning variables (panel A) these descriptive statistics are not necessarily informative about *causal* average programme effects due to possible selection biases. How to draw inference about the effects of these programmes is discussed in next section.

4 Econometrics

4.1 The causal modelling framework and the parameters of interest

We use Rubin's (1974) potential outcome language to describe a multiple treatment model under unconfoundedness, or conditional independence (Imbens, 2000, Lechner, 2001). Let D denote the treatment, which is non-participation and participation in one of the three programmes in our case. Thus, it takes on four different integer values from *0* to *3*. The (potential) outcome of interest that realises under treatment d is denoted by Y^d. For each individual, we observe only the particular potential outcome related to the treatment status that the individual has chosen, $y_i = \sum_{d=0}^{5} \underline{1}(d_i = d) y_i^d$ ($\underline{1}(\cdot)$ denotes the indicator function, which is one if its argument is true and zero otherwise).[6] There are two groups of variables to condition on, $\tilde{X}$ and Z. $\tilde{X}$ contains those covariates that are needed to correct for selection bias (confounders), while Z contains variables that define (groups of) population members for which an average causal effect estimate is desired. For identification, $\tilde{X}$ and Z may be discrete, continuous, or both, but for estimation, we will consider discrete Z only. They may overlap in any way. In line with the machine learning literature, we call them 'features' from now on. Denote the union of the two groups of variables by X, $X = \{\tilde{X}, Z\}$, $\dim(X) = p$.

Below, we investigate the following average causal effects:

$$IATE(m,l;x,\Delta) = E(Y^m - Y^l \mid X = x, D \in \Delta) \ ,$$

$$GATE(m,l;z,\Delta) = E(Y^m - Y^l \mid Z = z, D \in \Delta) = \int IATE(m,l;x,\Delta) f_{X|Z=z,D\in\Delta}(x)dx \ ,$$

$$ATE(m,l;\Delta) = E(Y^m - Y^l \mid D \in \Delta) = \int IATE(m,l;x,\Delta) f_{X|D\in\Delta}(x)dx \ .$$

---

[6] If not obvious otherwise, capital letters denote random variables, and small letter their values. Small values subscripted by '*i*' denote the value of the respective variable of individual '*i*'.

The **I**ndividualized **A**verage **T**reatment **E**ffects (IATEs), $IATE(m,l;x,\Delta)$, measure the mean impact of treatment *m* compared to treatment *l* for units with features *x* that belong to treatment groups Δ, where Δ denotes all treatments of interest. The IATEs represent the causal parameters at the finest aggregation level of the features available. On the other extreme, the **A**verage **T**reatment **E**ffects (ATEs) represent the population averages. If Δ relates to the population with *D*=*m*, then this is the **A**verage **T**reatment **E**ffect on the **T**reated (ATET) for treatment *m*. The ATE and ATET are the classical parameters investigated in many econometric causal studies. The **G**roup **A**verage **T**reatment **E**ffect (GATE) parameters are in between those two extremes with respect to their aggregation levels. The analyst preselects the variables *Z* prior to estimation according to her policy interest. The IATEs and the GATEs are special cases of the so-called **C**onditional **A**verage **T**reatment **E**ffects (CATEs).

4.2 Identification

The classical set of unconfoundedness assumptions consists of the following parts (see Imbens, 2000, Lechner 2001):

$$\{Y^0,Y^1,Y^2,Y^3\} \coprod D \mid X = x, \qquad \forall x \in \chi; \qquad (CIA)$$
$$0 < P(D = d \mid X = x) = p_d(x), \qquad \forall x \in \chi, \forall d \in \{0,...,3\}; \qquad (CS)$$
$$Y = \sum_{j=0}^{3} \underline{1}(D = j)Y^j; \qquad (SUTVA)$$

The conditional independence assumption (CIA) implies that there are no features other than *X* that jointly influence treatment and potential outcomes (for the values of *X* that are in the support of interest, χ). The common support (CS) assumption stipulates that for each value in χ, there must be the possibility to observe all treatments. The stable-unit-treatment-value assumption (SUTVA) implies that the observed value of the treatment does not depend on the treatment allocation of the other population members (ruling out spillover and treatment size effects). Usually, to have an interesting interpretation of the effects, it is required that *X* is not influenced by the treatment (exogeneity). If this set of assumption holds, then all IATEs are identified:

$$\begin{aligned}
IATE(m,l;x,\Delta) &= E(Y^m - Y^l \mid X = x, D \in \Delta) \\
&= E(Y^m - Y^l \mid X = x) \\
&= E(Y^m \mid X = x, D = m) - E(Y^l \mid X = x, D = l) \\
&= E(Y \mid X = x, D = m) - E(Y \mid X = x, D = l) \\
&= IATE(m,l;x); \qquad\qquad \forall x \in \chi, \forall m \neq l \in \{0,...,3\}.
\end{aligned}$$

Note that IATE does not depend on the conditioning treatment set, Δ. Since the distributions used for aggregation, $f_{X|Z=z,D\in\Delta}(x)$ and $f_{X|D\in\Delta}(x)$, relate to observable variables (*X*, *Z*, *D*) only, they are identified as well (under standard regularity conditions). This in turn implies that the GATE and ATE parameters are identified (their dependence on Δ remains, if the distribution of the features depends on Δ).

It is of course important that these conditions are plausible in our study. Let us consider them in turn. In Section 3 we already argued that availability of a wide range of socio-demographic information and of rich information about the labour market history of individuals enhances the plausibility of the CIA. These are essentially the variables identified by other evaluation studies as the most important confounders (e.g. Heckman et al., 1998; Lechner and Wunsch, 2013). These are also the variables available to the caseworker during the interview and thus should be the ones she is mainly basing her decision on. Advantages of our study compared to the training programme evaluation literature are the availability of sickness absence records as well as the unemployment rate in the district of residence. Probably the biggest disadvantage is the lack of earnings histories. However, this may not be so important as earnings are not an outcome variable (and thus earnings records are not needed for the role of pre-treatment outcomes) as well as because proxies for earnings are available, such as education, nationality, the sector of the previous jobs, the duration of the preceding employment spell, and the preferred desired profession of the job seeker. Overall, we conclude that CIA may be plausible. However, as a safeguard against possible violations we report a placebo study (that does not indicate any violations) below.

SUTVA is plausibly fulfilled as all programmes considered are rather small compared the labour force. Common support is a condition that can be checked in the data. We did not detect any common support problems with the programmes finally investigated. Finally, the exogeneity of confounding and heterogeneity variables is ensured by measuring all time varying variables at the beginning of the unemployment spell. At that moment, the individual did not know if and when she will enter a training programme.

4.3 Estimation

In this paper, we utilize the recently upcoming causal machine learning literature (see Athey 2019, and Athey and Imbens, 2019, for overviews). It combines the prediction power of the machine and statistical learning literature (for an overview see, e.g., Hastie, Tibshirani, and Friedman, 2009) with the microeconometric literature on defining and identifying causal effects (e.g., Imbens and Wooldridge, 2009). Recently, this literature has seen a surge of proposed methods, in particular in epidemiology and econometrics. Knaus, Lechner, and Strittmatter (2018) compare many of those methods systematically with respect to their set-up as well as their performance in a simulation exercise. One conclusion from their paper is that random forest-based estimation approaches outperform alternative estimators.

The starting point of the causal forest literature is the causal tree introduced in a paper by Athey and Imbens (2016). In a causal tree, the sample is splitted sequentially into smaller and smaller strata, in which the values of X become increasingly homogenous, to mitigate selection effects and to uncover effect heterogeneity. Once the splitting is terminated based on some stopping criterion, the treatment effect is computed within each stratum (called a 'leaf') by computing the difference of the mean outcomes of treated and controls (possibly weighted by the conditional on X probabilities of being a treated or control observation). However, the literature on regression trees acknowledges that the sample may be rather unstable because of its sequential nature (if the first split is different, the full tree will likely lead to different final strata). A solution to this problem is the so-called random forests estimator. Their key idea is to induce some randomness into the tree building process, build many trees, and then average the predictions of the many trees. The induced randomness is generated by using randomly generated subsamples (or bootstrap samples) and by considering for each splitting decision only a random selection of the covariates. Wager and Athey (2018) use this idea to propose causal forests, which are based on a collection of causal trees with small final leaves.[7] Lechner (2018) develops these ideas further by improving on the splitting rule for the individual trees, and by providing methods to estimate heterogeneous effects for a limited number of discrete policy variables (**G**roup **A**verage **T**reatment **E**ffects, GATE) at low computational costs, in addition to the highly disaggregated effects the literature focused on so far (**I**ndividualized **A**verage **T**reatment **E**ffects, IATE). Furthermore, Lechner (2018) suggests a way of performing unified inference for all aggregation levels. Finally, the approach is applicable to a multiple, discrete treatment framework. Since many of these advantages are important in the empirical analysis of this paper, this approach, termed Modified Causal Forests (MCF), is used below. For all further technical details of the estimator, the reader is referred to Lechner (2018).

4.4 Practical implementation

4.4.1 Outcome and control variables

We consider three types of outcome variables based on labour market history: employment, unemployment and a residual category which we call out-of-the-labour-force. It is de-

[7] Athey, Tibshirani, and Wager (2018) generalize this idea to many different econometric estimation problems.

fined as not being employed nor unemployed. These variables are either measured at a particular distance to the start of the programme, or in cumulated fashion as sum over a certain period.

The control variables have already been discussed in the previous sections. A complete list of them including descriptive statistics is contained in Appendix A. It is of course interesting to understand which of the features are important in the estimation. In classical programme evaluation of average population effects, such information would be deduced from an estimated propensity score. For Random Forest type estimators, computing so-called variable importance measures are informative about the relevance of one variable given all the others. They are computed by comparing the values of the objective function (estimated with out-of-bag observations, i.e. out-of-sample) of a prediction using the full set of variables with a prediction were the values of a specific variable are randomly permutated (so that this permutated variable becomes uninformative). In our case, the most important variables consisted of the country of birth, language skills, the simulated start date, labour market history (past employment and unemployment over various horizons), region, and the sector of the last employment. It is however important to note that a variable importance test will pick variables are either relevant for selection bias, or effect heterogeneity, or both. Those two aspects cannot be separated in a variable importance measure as they both determine the value of the objective function of the MCF.

4.4.2 Differential programme starts

Because individuals could be assigned to an ALMP at any point of time in their unemployment spell (although usually they are assigned in the beginning), we face a dynamic assignment problem. In such an environment the assumption of no anticipation is required in addition to the CIA and the construction of an appropriate comparison group is complicated, as first acknowledged by Fredriksson and Johansson (2008). No anticipation means that individuals do not alter their behaviour in response to a future assignment to the ALMP. Since in the period of analysis the training capacity tended to exceed demand, the time between assignment and the actual programme start is short, so that the bias induced by the failure of this assumption is likely to be small.

To transform a dynamic programme assignment into a static one, non-participants are defined to be the population that did not participate in the programme within a certain period, such as the first 9 months in this paper. Fredriksson and Johansson (2008) explain that such a definition biases the estimation of the effects upwards, as nonparticipants are less likely to have entered a programme, because they may have already found a job. To avoid this bias, they propose to define the comparison group as those that have not yet been treated. Based on these insight, two strands have developed in the literature. A first strand, aims at identifying the effects of those who did not *yet* receive a treatment (e.g. Sianesi 2004, 2008 and Biewen, Fitzenberger, Osikominu and Paul 2014). A disadvantage of this approach is that it redefines the effect and makes it dependent on the fraction of nonparticipants that participate (shortly) after this period.[8] Another strand of the literature therefore aims at identifying the effect relative to never receiving the treatment. This is essentially done by right censoring nonparticipants who subsequently enter the programme. Fredriksson and Johanson (2008) assume *independent* right censoring, while Crépon, Ferracci, Jolivet and van den Berg (2009) and Vikström (2017) generalize this by allowing for *selective* right censoring. Van den Berg and Vikström (2019) consider long-run post-treatment effects, such as those of vocational programmes on earnings.

Identifying the effects relative to never receiving the treatment with CML methods is beyond the scope of this paper. Here, we follow the first strand in this literature. We essentially follow the approach proposed by Lechner, Miquel and Wunsch (2011), which adapts the one suggested by Lechner (1999, 2002) to accommodate for the critique of Fredriksson and Johans-

8 In our empirical application about 25% of the nonparticipants enter an ALMP between 10 and 30 months after the beginning of the unemployment spell.

son (2008). Instead of regressing the log of the elapsed time to programme start within the unemployment spells of participants on a selection of the available explanatory variables that seem important for the timing of the programme, we use a post-LASSO estimator (i.e. OLS with the variables selected by LASSO estimation) to determine the relevant variables and the coefficients of this regression. We then use the estimated coefficients together with a draw from the residual distribution to predict the 'pseudo' programme starts for nonparticipants. Thus, the underlying assumption is that the assignment of programme start dates is random conditional on the variables included in the post-LASSO procedure. We exclude those nonparticipants for whom this simulated start date lies outside the 9-month treatment window and – to accommodate for the critique – those who are no longer unemployed at the assigned start date. The details of the determination of the pseudo programme starts can be found in Appendix B.

5 Results

In this section, we report the main results. We start by considering the average population effects for several outcomes of policy importance and their development over time. This informs us about the overall effectiveness of the different programmes and the dynamics of the effects. Next, we investigate whether the average population effects (ATE) differ from the effects of those unemployed workers in a particular programme (ATET). These comparisons are informative to understand the effects of caseworkers' selection to some extent. If caseworkers select programmes that are most effective for their specific unemployed, then ATET should be larger than ATE.

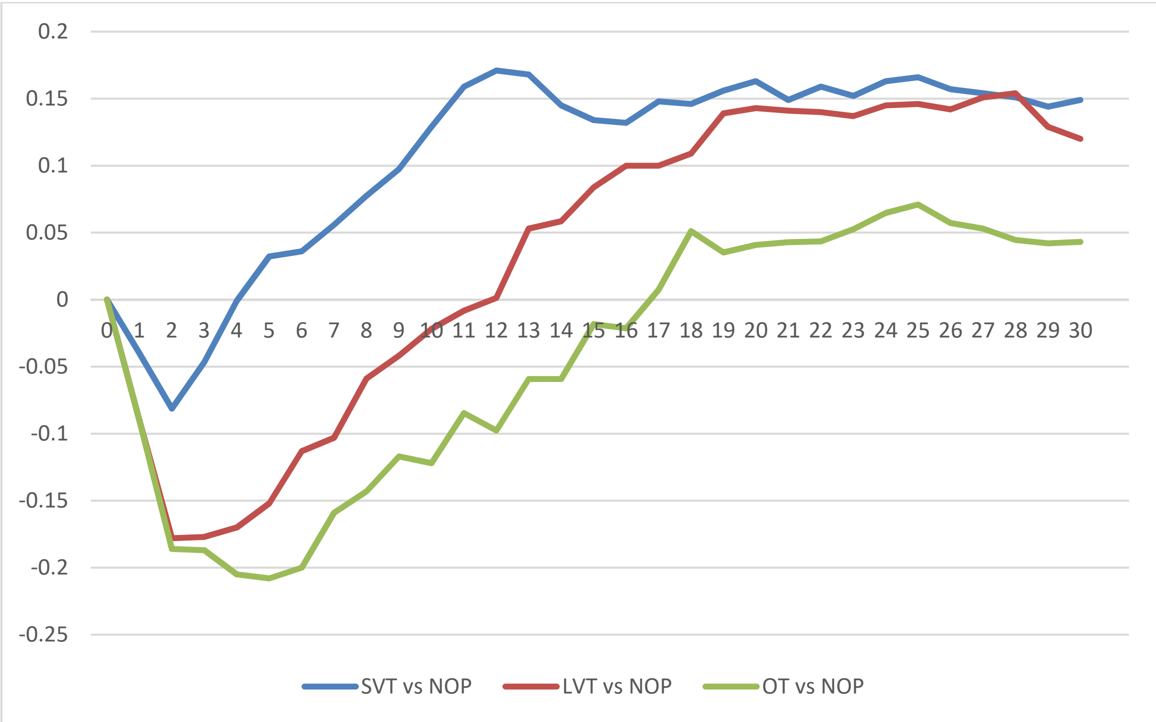
Then, for the arguably most important short- and medium-run outcome, namely employment, we investigate more thoroughly the heterogeneities with respect to the programmes and groups of unemployed by their programme participation. Subsequently, the heterogeneity of the most policy relevant medium-run effects are investigated with respect to a few variables considered to be of importance for the policy. Finally, in the last subsection we present an analysis of the IATEs, i.e. the effect estimates at the finest possible level of granularity.

5.1 Average population effects

5.1.1 Dynamics and programme heterogeneity

In this section, we report the average population effects (ATE) of the different programmes in comparison with no ALMP participation (NOP) and with each other.

Figure 5.1: The time evolution of the ATEs of the medium-run employment outcome



Note: Effects for month 1 are interpolated.

Figure 5.1 reports the dynamic evolution effects of the different programmes in comparison with no ALMP participation (NOP) on the probability to be employed.[9] Participation in short-term vocational training (SVT) modestly decreases the probability to be employed only during the first four months by a maximum of 8 percentage points (pp) relative to the counter-

[9] Standard errors and confidence intervals are omitted for clarity of presentation. All effects have a standard error of about 2.3-2.5 percentage points after month 20.

factual of NOP. Thereafter, the gain in employment is positive. The lock-in effect lasts about as long as the average programme duration of 3.8 months, which suggests that programme participation increases job-finding rates rapidly after the end of the programme. Subsequently, the ATE on the employment probability continues to rise until about 12 months after the programme start. Thereafter it stabilizes to around 15 pp, which is a substantial effect (statistically well determined), in particular if it remains stable over time as might be conjectured from its dynamic pattern.

Participation in long-term vocational training (LVT) leads to an employment probability which falls much more sharply during the first two months, reducing it up to 18 pp relative to the counterfactual of NOP. This follows naturally from the longer programme duration (7.2 months on average), but the lock-in period lasts even longer, until about one year after the programme start. The eventual impact of programme participation is of a similar magnitude as of SVT, i.e. around 15 pp., but it is only attained after about 20 months. This means that the longer time investment in human capital accumulation is not reflected in higher employment chances. It is possible that the higher time investment of LVT results in higher productivity and/or wage effects, but due to data unavailability this could not be tested. Thus, on average it appears that SVT dominates LVT as its courses are cheaper and its indirect costs (lock-in period) are lower as well.

The negative effects during the lock-in period of orientation training (OT) are even more pronounced as the effect in terms of the employment probability declines even to minus 21 pp, it takes 17 months before it becomes positive. This long lock-in effect is presumably related to 45% of OT participants entering other programmes within 3 months after completing OT. OT (including its follow-up programmes) is however also less effective in the medium run as its effect stabilizes around 5 pp, 10 pp below the level of VT. In conclusion, on average, SVT dominates LVT in terms of effectiveness, which in turn dominates OT.

Next, to get a better overall picture of the effects, we investigate (i) three summary measures of the employment effects (summed up over the first and last 9 months as well as over all 30 month), and (ii) two alternative outcome measure (months in unemployment, months out-of-the-labour-force).

Table 5.1: Effects for the different programmes 30 months after programme start on cumulative months in employment, unemployment and out of the labour force (ATE)

	No ALMP participation (NOP)	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
Cumulative months in *employment 9 months* after programme start				
NOP	**3.5 (0.0)**			
SVT	0.1 (0.2)	**3.6 (0.2)**		
LVT	-1.1 (0.1) ***	-1.2 (0.2) ***	**2.4 (0.1)**	
OT	-1.6 (0.1) ***	-1.6 (0.2) ***	-0.4 (0.2) **	**6.2 (0.2)**
Cumulative months in *employment between month 22 and month 30* month after programme start				
NOP	**5.7 (0.0)**			
SVT	1.4 (0.2) ***	**7.1 (0.2)**		
LVT	1.3 (0.2) ***	-0.1 (0.3)	**7.0 (0.2)**	
OT	0.4 (0.2) **	-0.9 (0.3) ***	-0.8 (0.3) ***	**6.2 (0.2)**
Cumulative months in *employment 30 months* after programme start				
NOP	**16.0 (0.1)**			
SVT	3.4 (0.5) ***	**19.4 (0.5)**		
LVT	1.0 (0.5) **	-2.4 (0.7) ***	**17.1 (0.5)**	
OT	-1.4 (0.5) ***	-4.8 (0.7) ***	-2.4 (0.7) ***	**14.7 (0.5)**
Cumulative months in *unemployment 30 months* after programme start				
NOP	**10.9 (0.1)**			
SVT	-1.9 (0.4) ***	**9.0 (0.3)**		
LVT	0.9 (0.4) **	2.8 (0.5) ***	**11.8 (0.4)**	
OT	2.7 (0.5) ***	4.5 (0.6) ***	1.8 (0.6) ***	**13.6 (0.5)**
Cumulative months out-of-the-labour force 30 months after programme start				
NOP	**3.1 (0.1)**			
SVT	-1.6 (0.3) ***	**1.6 (0.3)**		
LVT	-1.8 (0.3) ***	-0.4 (0.3)	**1.1 (0.3)**	
OT	-1.4 (0.3) ***	0.2 (0.4)	0.7 (0.4)	**1.7 (0.2)**

Note: Outcomes measured in months. Level of potential outcome for the specific programme on main diagonal in bold. All effects are population averages (ATE). Standard errors are in brackets. *, **, *** indicate the precision of the estimate by showing whether the p-value of a two-sided significance test is below 10%, 5%, 1% respectively.

The first panel of Table 5.1 shows that after 9 months SVT leads to the same number of months employment as NOP (3.5), while LVT and OT lead to substantial average losses of 1.1 and 1.6 months respectively. The cumulative effects in the last 9 months of the observation window (month 22 to 30) are all positive. They are largest and similar for SVT and LVT (1.3-1.4) and about one third of their magnitude for OT (0.4). These statistically well determined results are also confirmed when comparing the effects of the different programmes directly with each other. The third panel is summarizing these employment effects over all 30 months. While

SVT (3.4) and LVT (1.0) have positive effects, the effect OT is negative (-1.4) due to its large lock-in component. The last two panels of Table 5.1 report the average total impact of the different programmes on the time spent in unemployment (UE) and out-of-the-labour force (OLF) 30 months after the programme start. While all programmes reduce time in OLF by about one and a half month, only SVT decreases the time in UE as well (by 1.9 months). LVT increase UE by almost 1 month, while OT increases time in UE by almost 3 months. Again, these are at least partly repercussions of the differential lock-in effects.

These findings can be rationalized with ideas in both the economic and psychological literature. First, from an economic perspective we expect that participation in training reinforces labour force participation if the option value of participation is eventually positive, which is consistent with the findings reported in Figure 5.1. Furthermore, participating in training helps workers setting clearer professional targets, because counsellors typically have such targets in mind when assigning them to training programmes. The psychological literature typically finds that goal setting leads to more time and effort spent on job search (see e.g. van Hooft and Noordzij, 2009; Latham *et al.* 2018).

5.1.2 Programme group heterogeneity

While in Table 5.1 we investigated all effects for the population of unemployed, now we analyse how the effects differ for the different populations participating in the different programmes. One motivation for this perspective is that if caseworkers assign programmes according to their individual effectiveness, then we expect the effects for their own population (e.g. the effects of SVT for those participating in SVT) to be the largest. The detailed results in Table C.1 in Appendix C clearly show that this is not the case. In Table 5.2, we show formal statistical (Wald-) tests for the equality of the effects over the four populations. We see that out of 30 tests there is only one clear rejection, which is of course expected at conventional significance levels. However, the reason for this rejection is that the ATET is worse than the ATE.

Table 5.2: Wald test of equality of effects in all four treatment specific subpopulations

Outcome variable	SVT – NOP	LVT – NOP	OT – NOP	LVT – SVT	OT – SVT	OT – LVT
Cumulative months in employment 0-9 months after …	3.6	6.1	16.5***	0.7	1.6	2.7
Cumulative months in *employment 21-30 months* after …	1.9	2.8	0.3	0.9	0.6	1.1
Cumulative months in employment 0-30 months after …	3.2	3.2	2.5	0.6	0.4	0.1
Cumulative months in unemployment 0-30 months after …	4.6	3.0	2.8	0.1	0.2	0.1
Cumulative months in out-of-the-labour-force 0-30 months after	4.9	7.3*	6.7*	2.2	0.5	1.0

Note: Under the null of equality, the test statistic is distributed as $\chi^2(3)$.

This means that either the treatment effects are fairly homogeneous in these populations or that caseworkers' assignment to the different programmes is close to random. Below we will show that effects are clearly heterogeneous, so that we can conclude that caseworkers fail to assign the unemployed to those programmes from which they would benefit most.[10] In Section 6 we discuss the gains that the PES could make by improving the assignments of unemployed to the different programmes.

5.2 Heterogeneity with respect to policy relevant variables

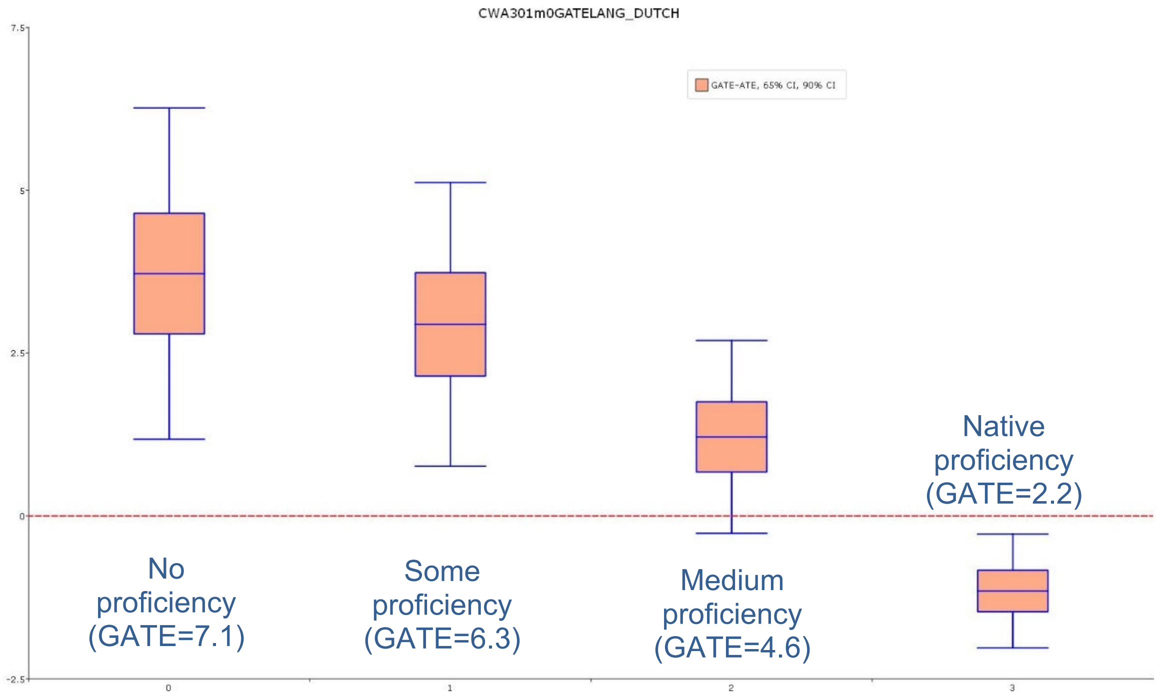
In many situations, there are heterogeneity variables a decision maker may particularly care about. In this section, we analyse such variables using the GATE parameter introduced above. We present the results for the overall population, as programme-population specific effects do not appear to deviate much from the population averages. We focus again on the main medium-term outcome: the cumulative number of months employed 30 months after the programme start. Of course, specifying a long list of policy relevant variables a priori and reporting significant results bears the danger of data-snooping. Here, we assume that the Flemish labour market authorities consider the following variables as particularly important: Unemployment history (last 2 and 10 years), unemployment duration at the start of the programme (below or above the median), age (younger than 25 or older than 50, below or above the median), sex,

[10] This appears to be in line with official policy. The PES in Flanders used outflow targets for evaluating caseworkers, like 70% of the participants must be employed within 6 months after the end of the training" instead of effect targets.

proficiency in Dutch language (4 point Likert scale/below highest proficiency level), unemployment rate in the district of residence at the start of the unemployment spell, country of birth (6 groups: Belgium, Southern EU countries, Eastern EU countries, other EU countries, Turkey or Morocco, rest of the world) , and 13 education levels (from second year of high school or below to master's degree).

Remember that we have sampled individuals who entered unemployment after being laid-off from a job. Recent unemployment history helps therefore identifying a population that is more loosely attached to the labour market in the sense that it can identify individuals who did not only experience employment in the past two years, but also some time in unemployment. From a policy perspective it could be interesting to identify programmes that work for such a population. A priori one could expect that the provision of vocational training may strengthen the competencies of this group and may accomplish more stable employment. It can, however, be useful to verify whether this hypothesis holds. This empirical evaluation based on a machine learning approach can help checking this and other hypotheses as the ones formulated below. By contrast, long-term unemployment histories can, for instance, help identifying a group of workers who had stable employment (by having little unemployment experience in the last 10 years), but who lost their job abruptly. This might be the group to which OT is targeted and it is of interest of knowing whether such a strategy works. In Belgium youth and older workers have difficulties in finding jobs, so that it is of interest to know which policies are effective for those younger than 25 or older than 50. Discrimination both in terms of gender and migration background (country of birth and proficiency in Dutch are proxies for this) is a very sensitive political issue in Flanders and in Belgium individuals with migration background have much more difficulty than elsewhere in the EU to find employment. Identifying which policies work best for containing such discrimination and for getting migrants to stable employment is therefore highly relevant. In the Belgian labour market low educated workers are particularly at risk of unemployment, so that knowledge about the relative effectiveness of policies according to the level of education of participants is valuable. Despite Flanders being a small region, unemployment rates vary substantially across districts. This is related to the limited geographic mobility within the region, induced, amongst others, by a policy that heavily supports home ownership and that stimulates traffic congestion. Finally, the effectiveness of training programmes according to the unemployment duration at which they start relates to the discussion of whether *preventive* or *curative* interventions are more effective.

Figure 5.2: Difference of GATEs to ATE of SVT relative to NOP for the four proficiency levels in Dutch – Cumulative number of months employed 30 months after programme start

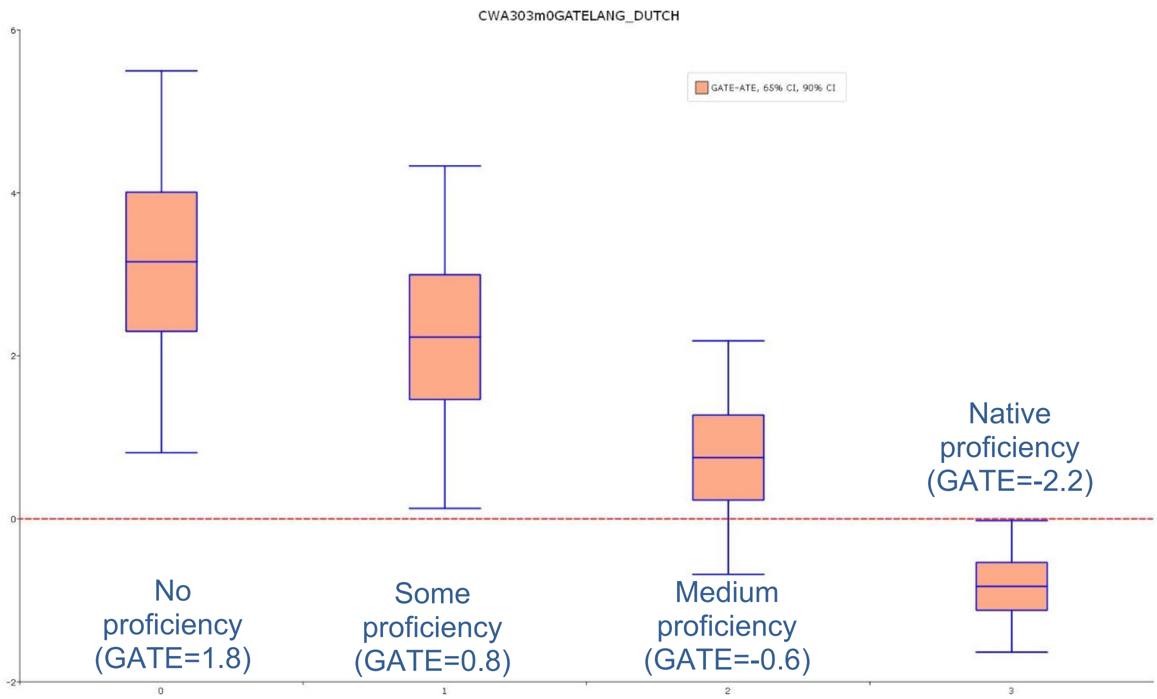


Note: Dutch proficiency varies between no proficiency (0) and native proficiency (3). The boxes indicate the 65% confidence intervals around the GATEs, while the ends of the whiskers mark the widths of the 90% confidence intervals. The vertical axis measures the deviation of the GATE from the ATE.

When we test for univariate treatment heterogeneity of the ATEs on the cumulative number of months employed 30 months after programme start, we find statistically significant differences at the 10% level in four dimensions (proficiency in Dutch, country of birth and level of education), but not for all three programmes. Figure 5.2 illustrates the difference of the GATEs (minus ATE) of SVT relative to NOP associated to the four proficiency levels by means of box-and-whisker plots. The horizontal line at zero indicates the level of the ATE. One can clearly observe a decrease in the GATEs with the proficiency level in Dutch and that the lower proficiency levels have significantly higher GATEs than the ATE. For instance, the GATE of those with no knowledge of Dutch is 3.7 months higher than the ATE (p-value of 2%).

Figure 5.3 reports the corresponding GATEs of OT versus NOP. Interestingly, the point estimates of the GATEs for the two lowest proficiency levels are positive while the point estimate of the ATE was negative. While these GATEs are not statistically significantly different from zero, they are statistically significantly different from the ATE, at the 3% level for proficiency level zero and at the 8% level for proficiency level one. The point estimates of the GATEs of LVT versus NOP also display a similar negative relationship with Dutch proficiency as reported for the other programme participations. However, none of these differences are significantly different from the ATE. We therefore do not display the corresponding figure.

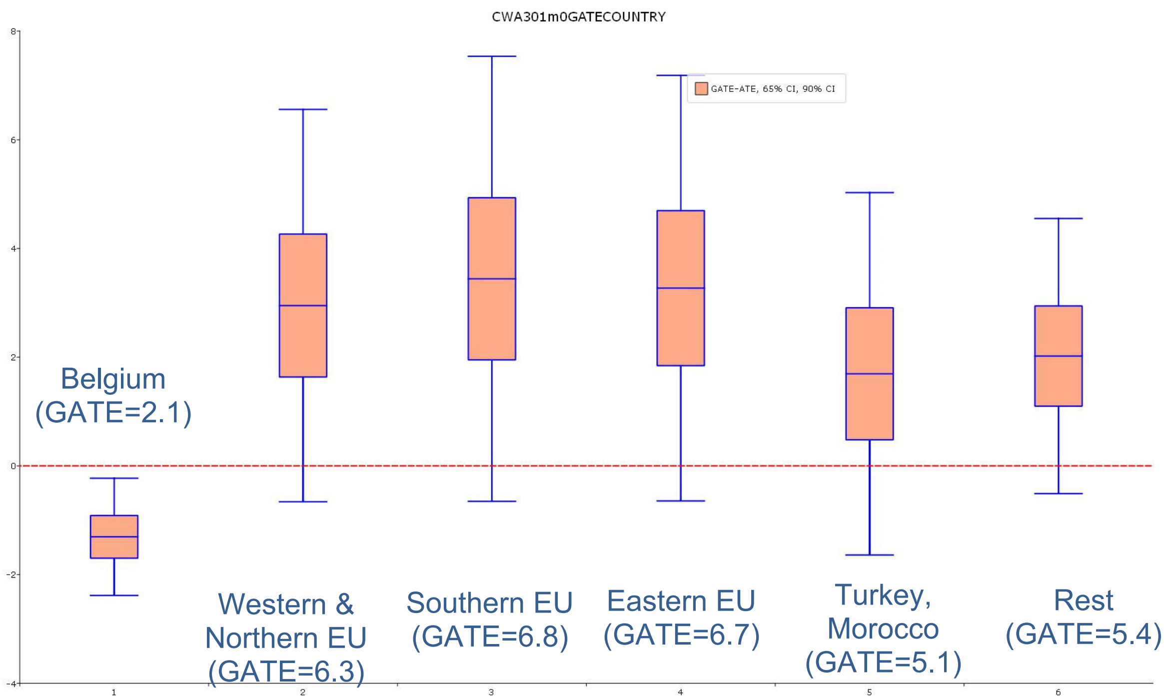
Figure 5.3: Difference of GATEs to ATE of OT relative to NOP for the four proficiency levels in Dutch – Cumulative number of months employed 30 months after programme start



Note: Dutch proficiency varies between no proficiency (0) and native proficiency (3). The boxes indicate the 65% confidence intervals around the GATEs, while the ends of the whiskers mark the widths of the 90% confidence intervals. The vertical axis measures the deviation of the GATE from the ATE.

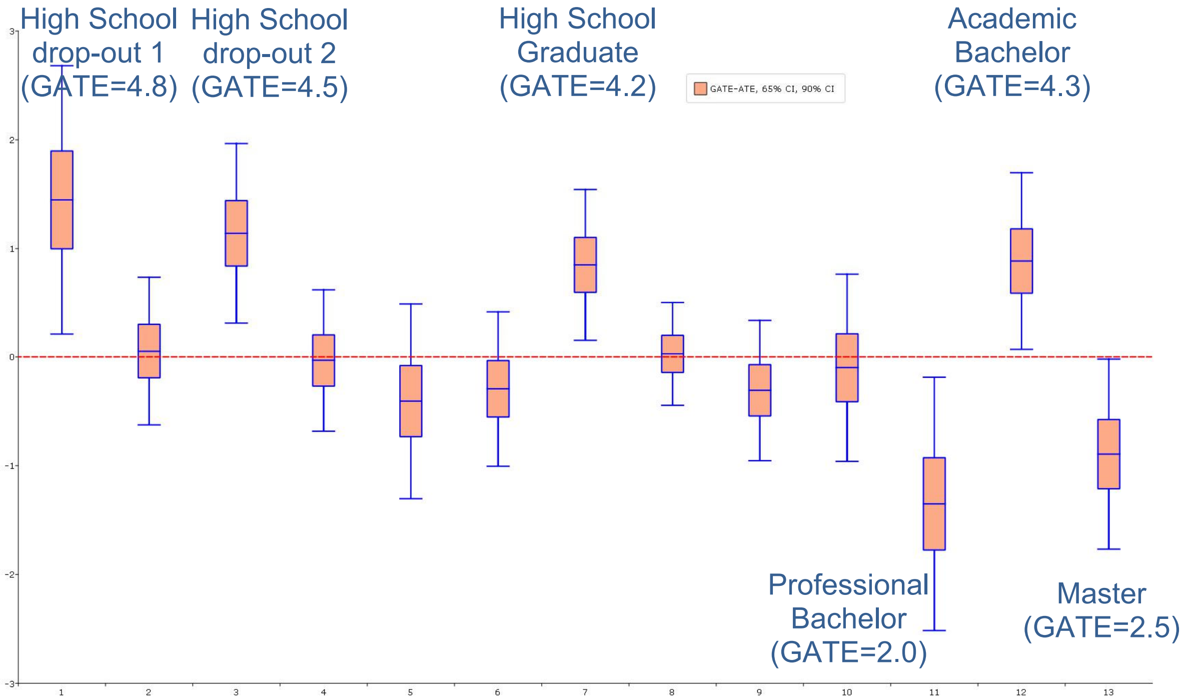
Figure 5.4 illustrates how the GATEs vary by country of birth. This suggests that the GATEs of SVT relative to NOP are the highest for individuals born in Southern European Union countries (6.8 months). It is notable that the effects for those born in Turkey and Morocco, i.e. for whom the employment rates are lower than for other foreigners, but they remain significantly higher (5.1 months) than for Belgians (2.5 months). Even if the precision is lower and we do not find statistically significant differences when considering the other ALMP, the pattern of the corresponding GATEs is similar and, hence, not reported. As additional evidence, we just compared the GATEs for being born in Belgium or not. For participants in SVT born outside of Belgium the GATE of SVT relative to NOP is 5.8 months compared to 2.1 months for those born in Belgium (p-value of 5%). Together with the previous finding this strongly suggests that SVT is more effective for migrants who recently migrated to Belgium.

Figure 5.4: Difference of GATEs to ATE of SVT relative to NOP according to country of birth – Cumulative number of months employed 30 months after programme start



Note: Country of birth on horizontal axes. Vertical axes denotes difference of respective GATE with ATE. The boxes indicate the 65% confidence intervals around GATE-ATE, while the ends of the whiskers mark the widths of the 90% confidence intervals. The vertical axis measures the deviation of the GATE from the ATE.

Figure 5.5: Difference of GATEs to ATE of SVT relative to NOP according to educational achievement – Cumulative number of months employed 30 months after programme start



Note: Final education level on horizontal axes. Vertical axes denotes difference of respective GATE with ATE. The boxes indicate the 65% confidence intervals around GATE-ATE, while the ends of the whiskers mark the widths of the 90% confidence intervals. The vertical axis measures the deviation of the GATE from the ATE.

Figure 5.5 displays the GATEs (minus ATE) of SVT relative to NOP for the 13 levels of education levels that we distinguish.[11] The highest effects are found for *general* education at different levels, but among these the GATEs are higher for lower levels of education, i.e., in decreasing order, for *first stage high school drop outs* who at most successfully completed the first qualification stage in high school (attained after the second out of six high school years), *second stage high school drop outs in the general track* who at most successfully completed the second qualification stage in high school (after four high school years), *academic bachelor*'s degree and *high school graduate in the general track*. The below average GATEs are found for *master*'s degree and, especially, for *professional bachelor's degree.* A similar ranking is

[11] Pupils can choose among four tracks in high school: (i) the general track provides a primarily theoretical general preparation for tertiary education; (ii) the technical track consists of a mix of theoretical and practical classes aimed at both direct labour market entry after completion and entry into primarily technical tertiary education; (iii) the vocational track teaches practical skills that prepare for particular professions; (iv) the arts track combines general education with active arts practice. The latter track is very small. In the analysis it is integrated in the general track.

found for the other programmes. Common to these results is that these programmes perform well for students who dropped out the educational process while they were following a general track without any practical professional orientation. This suggests that the vocational and orientation programmes are more effective for these groups because they provide training in specific professional skills that are directly useful on the labour market which these groups are lacking. Conversely, such programmes are less useful for those who already have accomplished some professional education that is directly useful on the labour market, and more so, the higher is the level of education that has been accomplished.

So far, we considered the GATEs for the cumulative number of months employed 2.5 years after the programme start. The heterogeneity in the effects of this outcome mixes two sources of heterogeneity: one during the lock-in phase and one during the post-treatment period. In both the study of Knaus, Lechner and Strittmatter (2017) and Bertrand, Crépon, Marguerie and Premand (2017) effect heterogeneity is essentially found during the lock-in phase and not so much post treatment. In our evaluation we confirm that heterogeneity is more important during this initial phase, but also find evidence of heterogeneity in the post-treatment effect.

During the lock-in phase TE heterogeneity is essentially caused by the differential speed at which different types of unemployed find employment in the potential state of NOP. So, generally this causes less negative programme effects for unemployed with a low employability, because even without programme participation these individuals would have low chances to transit to a job. To obtain an idea of the effect heterogeneity during the lock-in phase, we consider the GATEs for the cumulative number of months employed 9 months after the programme start. We do find evidence of heterogeneity in more dimensions than for the benchmark outcome. In addition to the aforementioned dimensions, all programmes are significantly *less* effective for youth below the age of 25 and *more* effective for older workers above age 50 with the effectiveness generally increasing with age, *more* effective for those living in a city, *more* effective for longer term unemployed (except for SVT), *less* effective for those with a lot of unemployment experience in the last two years.

To evaluate the heterogeneity in the post-treatment, we consider the cumulative employment outcome between months 22 and 30 after the programme start. This is when the programme effects are in their long-run equilibrium (see Figure 5.1). In this case the heterogeneity remains in the following dimensions: higher effectiveness for those with lower level of Dutch proficiency, for those born in a foreign country, in particular those born in a Southern or Eastern EU country. For education the point estimates display similar differences as for the benchmark outcome, but they are no longer statistically significant.[12]

5.3 Heterogeneity at the (averaged) individual level (IATEs)

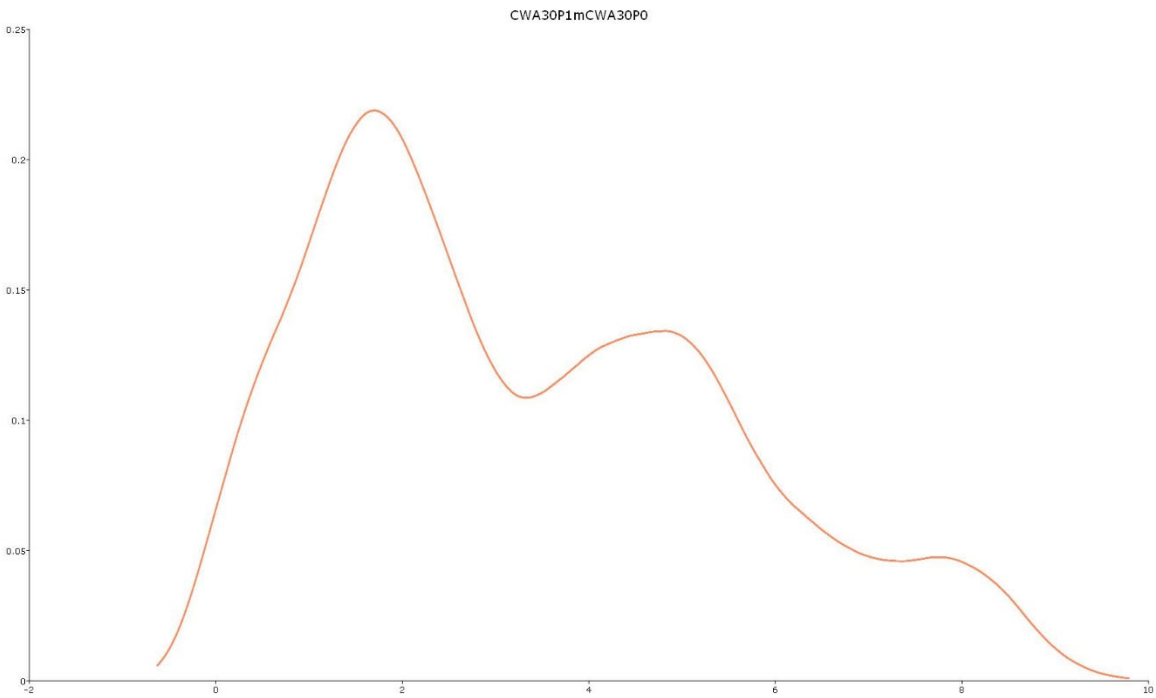
In this section, we present the results for the individualized average effects (IATEs), which present the finest level of granularity available. To avoid flooding the reader with numbers, we will concentrate on the cumulative medium-term employment outcomes for the comparison to NOP (no programme participation) which are likely to be the most policy relevant. We first describe the extent of heterogeneity in the programme effects. We then report the results of a post LASSO regression of the IATEs on the conditioning variables to detect systematic partial relationships between the former and the latter. Finally, we present the results of a k-means clustering analysis to get an informal characterization of sub-groups according to the effectiveness of programme participation.

Figure 5.6 shows the distribution of the IATEs of SVT vs. NOP. 99% of the estimated effects are positive, have a mean of 3.4 month (as shown in Table 5.1) and a standard deviation of 2.2. With a median standard error of the estimates of the IATEs of about 2.1 about 34% of

[12] Detailed results on lock-in as well as medium run heterogeneity are available on request from the authors.

the estimated IATEs are significantly different from zero. This points to two important issues: (i) There is considerable heterogeneity in the IATEs, some of which however is due to estimation error; (ii) It is much more difficult to get a precise estimate (without imposing functional forms) for the IATEs than for the GATEs and ATE that were estimated with rather high precision. These features are also visible when considering Figure 5.7, in which the sorted effects are given together with a 90%-confidence interval based on the estimated standard errors (see also Chernozhukov, Fernandez-Val, and Luo 2018). Again, we see a substantial variation of the effects, but also that the uncertainty of the ATE is much lower than for the IATEs.

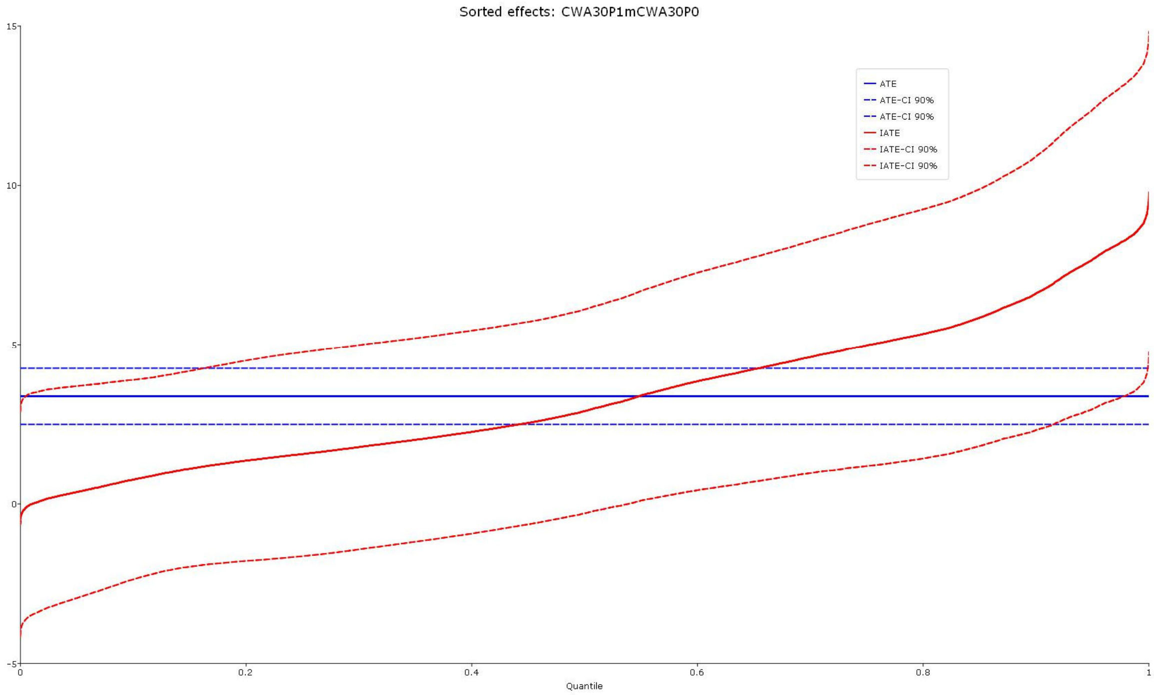
Figure 5.6: Distribution of estimated IATE of SVT vs. NOP



Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.

The respective figures for the other programmes are like the ones presented and are moved to Appendix C.2.

Figure 5.7: Overall heterogeneity: sorted effects of SVT relative to NOP – Employment 30 month after the programme start



Note: IATEs are sorted according their size. 90%-confidence interval based on estimated standard errors and normal distribution. Standard errors are smoothed by Nadaraya-Watson regression (Epanechnikov kernel with Silverman bandwidth).

From the previous discussion of the GATEs and the distribution of the IATEs there is substantial effect heterogeneity. In the previous section we discussed to what extent this heterogeneity is directly related to policy relevant variables. However, by considering the marginal effects of one variable at a time, some of the heterogeneity may not be detected, because they may be confounded with other explanatory variables. However, to be able to detect further patterns of heterogeneity at this fine level is impossible without some additional structure (due to high estimation noise otherwise). Therefore, we use first a linear model estimated by Post-LASSO and second a k-means clustering algorithm to understand further possible heterogeneities better.

To implement the Post-LASSO procedure, we randomly split the sample into two. In first part we run a LASSO. In the second part we run OLS using the variables with non-zero coefficients in the LASSO estimation.[13] That way, standard OLS inference remains valid (given the selected model and the estimated dependent variables).

A first observation is that the LASSO selects many variables: As many as 124 out of 175 variables for the comparison SVT to NOP and 122 variables for the other two comparisons to NOP. Moreover, most of the estimated coefficients are statistically significantly different from zero. Most GATE results are confirmed with some nuance. Proficiency in Dutch and born in Belgium decreases effectiveness. However, there is no specific region of origin that uniformly performs better: This depends on the considered programme. Higher education, except for academic bachelors,[14] generally decreases the effectiveness. However, the programmes no longer seem most effective for high school drop-outs, but rather for those who completed part-time professional education and special education for persons with some physical or mental disability.[15] High school dropout seems to be confounded by other factors that determine programme effectiveness.

The Post-LASSO regressions indeed reveal some new dimensions in which the programme effectiveness varies: effectiveness consistently increases for women, with age, with the unemployment duration at which the programmes start. The finding of higher effect for women and long-term unemployed is in line with the results of meta-analysis (Card, Kluve and Weber, 2018). In addition, recent (in the last 2-5 years) experience in both unemployment and employment, decreases programme effectiveness. This means that programmes are more effective for those who were out of the labour force, possibly because they were not living in the country. For those with high long-term (within last 10 years) employment history the programmes seem to work better, presumably because their employment experience became ob-

[13] Categorial variables are expanded into a full set of dummy variables.

[14] This exception does not hold for OT.

[15] Since individuals who are currently disabled are excluded from the dataset, these must be individuals with some mild or temporary disability.

solete. We also find that for those who entered the labour market for the first time as unemployed or for those who at some point since 1991 were recipients of social assistance the programmes systematically perform worse. There are further dimensions in which heterogeneity is detected, such as the previous sector of employment, the type of and experience in the desired profession, having kids, the region of residence and whether one lives in a city or not. However, for these dimensions the direction of the effectiveness depends on the programme.

Next, complementing the heterogeneity analysis of the previous section, we describe the dependence of the effects on covariates by *k-means++* clustering (Arthur and Vassilvitskii 2007). The clustering is implemented by jointly using the IATEs of the 3 programme effects relative to NOP to form eight clusters. For reasons of conciseness, the clustering is only presented for the cumulative employment outcome 30 months after the programme start. The results are contained in Table 5.3.

The clustering is close to uniformly monotone in the effectiveness of all programmes and the columns in Table 5.3 are ordered accordingly. The analysis reveals again the important heterogeneity in the programme effects. The employment gains range from 1.2 to 8.0 months for SVT, from -1.7 to +4.9 months for LVT and from -2.8 to +1.9 for OT. The programmes are clearly the most effective for those with the lowest proficiency in Dutch, born abroad and with very little recent unemployment *and* employment experience, on average respectively one and five months in the last two years. This profile can only match recent entries into the labour market. Eastern and Southern EU are the most representative countries of origin of the most effective group, but it is notable that individuals from of Turkey and Morocco and from the rest if the world – migrants with worst labour market performance in Flanders[16] – are most represented in the second most effective group. Taken this together (and that the mean age rules out

[16] In 2018 the employment rate for those born in a non-EU28 country was 61.2% while it was 76.2% for natives (Source: Labour force survey as reported by www.steunpuntwerk.be).

recent school leavers), it become obvious that the group for which the programmes are most effective consist mainly of recent migrant. Additionally, living in a city rather than in rural areas and postponed programme starts are also associated with larger programme effects.

The two least effective groups are natives with excellent proficiency in Dutch and relatively much recent (last 2 years) and less recent (last 10 years) employment experience. Their first entry on the labour market was typically as unemployed job seekers. It is also notable that there is no clear relationship between programme effectiveness on the one hand and age and gender on the other.

Table 5.3: Descriptive statistics of clusters based on k-means clustering

Cluster	**Least beneficial**	**2**	**3**	**4**	**5**	**6**	**7**	**Most beneficial**
Share of observations in %	**20**	16	18	16	10	11	5	**5**
	Mean							
	Individualized average treatment effects (IATE) for the comparison to NOP (no participation)							
SVT-NOP	**1.2**	1.4	2.4	4.2	5.0	6.0	6.0	**8.0**
LVT-NOP	**-0.1**	-1.7	0.5	1.1	2.7	3.9	2.1	**4.9**
OT-NOP	**-2.6**	-2.8	-1.7	0.0	-2.8	-0.6	2.5	**1.9**
	Selected features							
Age	**32**	28	40	39	36	36	37	**35**
Women (in %)	**74**	0	60	40	61	65	12	**48**
Living in a city (in %)	**29**	32	29	28	51	45	48	**47**
Proficiency in Dutch (3: high, 0: none)	**3.0**	3.0	2.8	1.9	2.4	2.3	1.1	**0.8**
Country of birth: Belgium (in %)	**100**	100	95	72	1	0	9	**0**
Country of birth: Western & Northern EU (in %)	**0**	0	0	1	12	38	12	**14**
Country of birth: Southern EU (in %)	**0**	0	0	1	0	4	5	**18**
Country of birth: Eastern EU (in %)	**0**	0	0	0	1	24	4	**58**
Country of birth: Turkey & Morocco (in %)	**0**	0	0	6	0	16	42	**7**
Country of birth: Rest of the World (in %)	**0**	0	5	18	86	18	28	**3**
BIT (unemployed at first labour market entry; in %)	**52**	72	23	18	27	16	5	**0**
# of months unemployed in last *10 years* prior to UI	**14**	19	24	18	29	15	13	**1.6**
# of months employed in last *10 years* prior to UI	**57**	49	56	60	52	37	23	**7**
# of months unemployed in last *2 years* prior to UI	**4**	4	5	4	5	3	3	**1**
# of months employed in last *2 years* prior to UI	**18**	17	16	17	17	14	10	**5**
# of days until programme start	**75**	84	109	96	97	106	98	**108**
Predicted outcome without programme (NOP)	**19**	18	17	15	13	13	13	**13**

Note: Outcome variable is cumulative employment 30 months after programme start. All IATEs for all comparisons to nonparticipation are used to form the eight clusters. Covariates are not used to form clusters. K-means ++ algorithm used (Vassilvitskii, 2007).

Finally, one way to characterise the employability of the unemployed is to consider their estimated months of employment under NOP. The last row in Table 5.3 shows that in those eight groups, programme effectiveness monotonically decreases with increasing employability, which is consistent with the picture of heterogeneity revealed so far.

6 Policy experiments

So far, we have documented considerable heterogeneity in the effectiveness of the programmes that we have evaluated. A natural question that then emerges is whether caseworkers exploit this information and assign the unemployed to the programmes that work best for them and, if not, to what extent a different assignment could improve the performance of the PES. We take up this question in this section by simulating some hypothetical programme allocations which we compare to the observed allocation.

Table 6.1: Overall effects of some simulated hypothetical programme allocations

	Share of different programme states in %				Cumulative number of months 30 months after programme start		
	NOP	SVT	LVT	OT	Em-ployed	Out-of-labour-force	Unem-ployed
Observed	94.0	2.1	2.0	1.8	16.10	3.01	10.89
Random	94.0	2.0	1.9	1.9	16.11	3.01	10.88
Optimal – no constraint	2.6	97.3	0.2	0.2	19.42	1.64	8.94
Optimal – no constraint, only significant	39.7	58.1	1.6	0.5	18.82	1.79	9.39
Optimal – constraint, preference to large gains	94.0	2.1	2.0	1.8	16.41	2.80	10.79
Optimal – constraint, preference to worst NOP outcomes	94.0	2.1	2.0	1.8	16.26	2.90	10.84
Optimal – constraint, preference to lots of UE in the past*)	94.0	2.1	2.0	1.8	16.21	2.94	10.85

Note: Allocations minimize unemployment and maximise employment (equally weighted). *) If programmes capacity becomes a constraint, preference is given to unemployment the highest number of months in unemployment over the last 10 years.

The first line of Table 6.1 describes the consequences in terms of cumulative number of months employed, unemployed (UE) and out of the labour force (OLF) 30 months after the programme start of the *observed* allocation to the different ALMP. In the first four columns the share of the total population allocated to the different programmes is reported, while the three last columns display the impact on these outcomes. The second line in this table shows the outcome of a simulation in which all members of the population would have been randomly assigned to the different programmes with probabilities of assignment equal to the observed shares of participation in the different programmes. This would in fact correspond to an allocation in which *no information* about the programme effectiveness would be used. The outcome of this simulation shows that the average programme impacts of the observed allocation are identical to the ones obtained with random assignment.

The previous result clearly shows that the Flemish PES can improve the performance of the ALMP by adjusting the assignment according to the expected programme performance of individuals as estimated by their IATE. We consider several alternative assignment schemes depending on the available programme capacity, the degree of certainty about the effectiveness of programme participation and the objectives of the PES. The third line shows the results in case that the PES would not have any capacity constraints and the PES would maximize the number of months in employment and minimize the number of months in UE, each with equal weight. In this case, the PES would allocate more than 97% of the unemployed to SVT, about 0.2% to LVT and to OT, and less than 3% would remain untrained. Such an assignment could on average increase the number of months in employment by 3.3 (an increase of about 20%) by reducing the number of months in UE by 1.9 and OLF by 1.4. Obviously, this would be very costly, because it would massively increase the number of SVT participants. Whether the benefits for society would outweigh these costs is unclear, since we do not have access to the required information to conduct such cost-benefit analysis.

The previous allocation ignores the fact that same estimated IATE are positive (or reverse) just because of estimation error. Therefore, in line four we report the outcome of a simulation in which we assign individuals only to programmes to the extent that the corresponding IATEs are significantly positive or negative at the 2.5% level of a one-sided statistical test, respectively in terms of their effect on the time spent in employment or UE. From this simulation we see that for a large part of the population (39%) the IATEs are not significantly different from zero and relatively small, as the programme effects decrease in a much lesser proportion. The average gain in terms of employment is still 2.7 months which is nearly 17%.

The next lines consider cases in which it is assumed that the training capacity of the PES is constrained to the current one (and thus programme costs remain the same). Since only 6% of the population participates in training this cannot but dramatically reduce the gains that a relocation can generate. Three scenarios are considered. In the first scenario, priority is given to individuals with the highest returns to programme participation.[17] The average employment gain for the population is still 9.3 days (i.e. 0.31 months). Per participant this is on average more than 5 months (or 155 days). This is indeed a substantial gain. In the second scenario we give priority to individuals who have the worst NOP outcome. We find that this assignment rule realizes about half of the potential gain (as has already been suggested by the clustering results of Table 5.3). This suggests that there is some trade-off between equity and efficiency if the relevant outcome measure is the cumulative number of months employed. Finally, when priority is given to those who have lots of unemployment experience in the past, we obtain again a slightly smaller gain. In conclusion, while the gains in reallocation could be substantial, simple allocation rules are not able to achieve them.

7 Sensitivity analysis

7.1 Placebo analysis

To further convince the reader (and us) that the matching variables sustain the CIA, we provide a placebo validation test like the one proposed by Imbens and Wooldridge (2009, pp. 48–50). This validation consists in estimating with the same methodology the ATEs within a

[17] In case of excess demand for programs, priority is given to those individuals for whom the difference in performance between the best and the second-best program is maximal. Such a rule performs better than assigning individuals to their best choice, because this avoids that the gain of doing so is destroyed by the loss that a second choice for rationed individuals imposes.

preceding unemployment spell of participating in a future training programme. Since (unanticipated) future participation in training should not have any impact on the current outcomes, finding an effect close to zero provides some support for CIA.

To implement this placebo test, we select from the population of analysis the subpopulation that has experienced at least one unemployment spell – in case of multiple spells, we retain the first observed one – starting between September 2008 and February 2014. February 2014 is used, because it leads to a gap of 9 months between the start of the last unemployment spell that was retained for the placebo sample and the first considered entry in the main analysis, i.e. December 2014. This gap allows to estimate the placebo treatment effects during 9 months since the start of the preceding unemployment spell. This choice of 9 months aims at finding a balance between not reducing the size of the placebo population too much – this size declines rapidly with the size of the gap time – and having a sufficiently long period over which to measure the placebo effects. To avoid contamination, we dropped all individuals who entered an ALMP during this preceding unemployment spell. The eventual sample on which this placebo analysis is conducted consists of 17,943 non-participants, and 360, 336 and 285 participants in SVT, LVT and OT, respectively.

Table 7.1 reports the results for three outcomes: cumulative number of months employed, unemployed and out-of-the-labour force 9 months after entry in the preceding unemployment spell. The results show that all ATEs are close to zero and precisely estimated despite the rather small programme groups.

Table 7.1: Placebo Effects for the different future programmes on cumulative months in employment, unemployment and out of the labour force (ATE)

	No ALMP participation (NOP)	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
Cumulative months in employment 9 months after entry in the preceding unemployment spell				
NOP	**3.9 (0.1)**			
SVT	0.01 (0.3)	**3.9 (0.3)**		
LVT	0.5 (0.3)	0.5 (0.4)	**4.3 (0.3)**	
OT	0.001 (0.4)	-0.02 (0.4)	0.5 (0.4)	**3.9 (0.2)**
Cumulative months in unemployment 9 months after entry in the preceding unemployment spell				
NOP	**4.8 (0.04)**			
SVT	-0.1 (0.3)	**4.8 (0.3)**		
LVT	-0.4 (0.3)	-0.5 (0.4)	**4.4 (0.3)**	
OT	-0.002 (0.3)	-0.1 (0.4)	0.4 (0.4)	**4.8 (0.3)**
Cumulative months out of the labour force 9 months after entry in the preceding unemployment spell				
NOP	**0.4 (0.01)**			
SVT	-0.1 (0.1)	**0.3 (0.02)**		
LVT	-0.1 (0.1)	-0.005 (0.1)	**0.3 (0.1)**	
OT	-0.03 (0.1)	0.04 (0.2)	0.04 (0.2)	**0.3 (0.1)**

Note: Outcomes measured in months. Level of potential outcome for the particular programme on main diagonal in bold. All effects are population averages for the respective placebo programme participants given in the column. Standard errors are in brackets. *, **, *** indicate the precision of the estimate by showing whether the p-value of a two-sided significance test is below 10%, 5%, 1% respectively.

7.2 Tuning parameters

To investigate the stability of the MCF estimates with respect to various tuning parameters (see also Appendix B for more details), we performed the following sensitivity exercises: (i) The number of bootstrap replications has been increased from 1000 to 2000 replications; (ii) the minimum leaf size has been varied from 5 to 3 and 7; (iii) the subsampling share was decreased from 67% to 50%; (iv) the number of variables used for splitting any particular leaf has been varied; (v) estimation was performed with and without prior deselection of irrelevant features, and (vi) the penalty term in the MCF objective function has been increased 10 fold from its base value that equals the variance of the respective outcome variable. None of these variations led to any substantial changes in the estimation results.

7.3 Distribution of weights

As it has already mentioned that estimated causal effects from Causal Forests have a representation as weighted means of the outcome variable. These weights can be investigated to check the stability of the estimation. If few weights are very large, this indicates that very few observations play a very important role to estimate the counterfactual. For example, Huber, Lechner, Wunsch (2013) considered weights with values larger 4% of the total (absolute) sum of weights as being a concern. It turned out that for the ATEs and the GATEs none of the weights are above 1%, respectively 3%. The exception is the GATE for the country of origin, for which in the subsamples of training participants, about 0.4% of the observations have weights between 4% and 10%. For the IATEs the situation is more extreme (as is reflected in larger standard errors shown above) as the weights are naturally more concentrated: In the training subsamples, about 1-2% of the weights are above 4% and in very rare occasions they could become as large as 25%. Further research will show the implications of such large weights, and how they might be adjusted to avoid small sample issue. In the same vein it becomes clear that much larger samples are needed for a reliable estimation of the IATEs than the for ATE or GATEs.

8 Conclusion

In this paper we used recent developments in causal machine learning to investigate the average and heterogeneous effects of very recent training programmes in Belgium, using administrative individual data from the Public Employment Service of Flanders. We found that on average all programmes have positive employment effects in the medium run, although sometimes not large enough to compensate, after 2.5 years, for the early negative effects in the lock-period. It turned out that on average short vocational training is more effective than longer vocational training courses as well as orientation training. Analysing the heterogeneity of the effects, the striking result appeared that programmes seem to work better (even after the lock-in period) for unemployed with a low employability, in particular recent migrants with limited language skills. Using the fine-grained results for analysing the assignment policy of Flanders' public employment service revealed considerable inefficiency. A different allocation of unemployed to existing programme slots should lead to a substantial improvement in labour market performance at no or small additional costs.

We may compare these findings with the meta study of Card, Kluve and Weber (2018) who analysed the effectiveness of active labour market policies (ALMP) based on more than 200 papers. In line with their general findings, we detect close to zero effects in the short-run due to lock-in, but that training programmes become effective after two to three years. They generally find that programmes with more human capital accumulation (i.e. training) are more effective in the longer run. Our findings nuance these conclusions as our evidence shows that SVT is as effective as LVT even in the long-run. Card et al. (2018) find that heterogeneity is relevant, but they do not report, as we do, higher effectiveness for (recent) migrants or with respect to the level of education. They report evidence of higher impacts for women, long-term unemployed and during recessions. We never find a differential impact for residence in high unemployment regions. Unemployment duration matters in the lock-in phase (9 months after programme start), aside of other factors that are negatively related with the employability in the absence of programme participation: the negative impact of lock-in diminishes as non-participants are less likely to be employed. We also find evidence that programmes are more effective for women, aside of many other factors, but only after controlling for other determinants of the programme effects in post-LASSO regressions in which we regress estimated effects on potential heterogeneity factors.

On a methodological note, the Modified Causal Forest (MCF) approach (Lechner, 2018) seems to be well suited for such an analysis and appeared to led to plausible and informative results at reasonable computational costs.

Future work could address many open issues, such as extending the data base with respect to additional control variable such that the effects of other programmes of the active labour market policy of Flanders that are ignored here can be credibly evaluated as well. Furthermore, it will be interesting to see whether similar heterogeneity appears in other countries with comparable policies. More generally, extending the CML framework such as to be able to address assignment dynamics is likely to lead to additional insight about policy effectiveness and assignment optimality. Finally, using an explicit optimal policy approach (e.g., Athey and Wager, 2019b, Zhou, Athey, and Wager, 2018) and comparing it with our IATEs based approach is an interesting exercise that could lead to further valuable insights about possible gains from alternative allocations of unemployed to programmes.

References

Arthur, D., Vassilvitskii, S. (2007): k-means++: the advantages of careful seeding, *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 1027–1035.

Athey, S. (2019): The Impact of Machine Learning on Economics, in Agrawal, A., J. Gans, and A. Goldfarb (eds.), The Economics of Artificial Intelligence: An Agenda. 507-547, Chicago: Chicago University Press.

Athey, S., and G. W. Imbens (2016): Recursive Partitioning for Heterogeneous Causal Effects, *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7353–7360.

Athey, S., and G. W. Imbens (2019): Machine Learning Methods Economists Should Know About, *arXiv*.

Athey, S., and S. Wager (2019a): Estimating Treatment Effects with Causal Forests: An Application, forthcoming in *Observational Studies*.

Athey, S., and S. Wager (2019b): Efficient Policy Learning, *arXiv*.

Athey, S., J. Tibshirani, and S. Wager (2019): Generalized Random Forests, *The Annals of Statistics*, 47, 1148-1178.

Autor, D. H., F. Levy, and R. J. Murnane (2003): The Skill-Content of Recent Technological Change: An Empirical Investigation, *Quarterly Journal of Economics*, 118, 1279-1333.

Bertrand, M., B. Crépon, A. Marguerie and P. Premand (2017): Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire, *mimeo,* University of Chicago, Crest and World Bank.

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Paul (2014): The effectiveness of public sponsored training revisited: the importance of data and methodological choices, *Journal of Labor Economics*. 32 (4), 837–897.

Card, D., J. Kluve and A. Weber (2018): What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations, *Journal of the European Economic Association,* 16(3). 894-934.

Chernozhukov, V., I. Fernandez-Val, and Y. Luo (2018): The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages, *Econometrica,* 86, 1911-1938.

Chou, Philip A. (1991): Optimal partitioning for Classification and Regression Trees, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13-4, 340-354.

Crépon, B., Ferracci, M., Jolivet, G. van den Berg, G.J. (2009): Active labor market policy effects in a dynamic setting, *Journal of the European Economic Association,* 7, 595–605.

Faltings, R., A. Krumer, and M. Lechner (2019): Rot-Jaune-Verde. Language and Favoritism: Evidence from Swiss Soccer, University of St. Gallen discussion paper, 2019-15.

Fredriksson P., and P. Johansson (2008): Dynamic treatment assignment: the consequences for evaluations using observational data, *Journal of Business and Economic Statistics,* 26(4), 435–445.

Friedman, J., Hastie, T. and Tibshirani, R. (2008): Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software*, 33(1), 1-22. Goos, M., A. Manning and A. Salomons (2009): The Polarization of the European Labor Market, *American Economic Review Papers and Proceedings* 99, 58-63.

Goos, M., A. Manning, and A. Salomons (2009): Job Polarization in Europe, *American Economic Review, P&P*, 99:2, 58-63.

Hastie, T., R. Tibshirani, and J. Friedman (2009): The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, Springer (10th printing with corrections, 2013).

Heckman, J. J., H. Ichimura, J. A. Smith, and P. E. Todd (1998): Characterizing selection bias using experimental data, *Econometrica* 66, 1017–1098.

Huber, M., M. Lechner, and C. Wunsch (2013): The performance of estimators based on the propensity score, *Journal of Econometrics*, 175 (1), 1-21.

Imbens, G. W. (2000): The Role of the Propensity Score in Estimating Dose-Response Functions, *Biometrika*, 87, 706-710.

Imbens, G. W., and J. M. Wooldridge (2009): Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, 47 (1), 5-86.

Knaus, M. C., M. Lechner, and A. Strittmatter (2017): Heterogeneous Employment Effects of Job Search Programs: A Machine Learning Approach, *arXiv: 1709.10279v2*.

Knaus, M. C., M. Lechner, and A. Strittmatter (2018): The Performance of Machine Learning Approaches to Estimate Heterogeneous Policy Effects, *arXiv: 1810.13237v1*.

Langenbucher, K. (2015): How demanding are eligibility criteria for unemployment benefits, quantitative indicators for OECD and EU countries, *OECD Social, Employment and Migration Working Papers*, No. 166, OECD Publishing, Paris, doi: 10.1787/5jrxtk1zw8f2-en.

Latham, G. P., M. B. Mawritz and E. A. Locke (2018): Goal Setting and Control Theory: Implications for Job Search, in: Klehe U.-C. and E. van Hooft (eds.), *The Oxford Handbook of Job Loss and Job Search*, Chapter 8.

Lechner, M. (1999): Earnings and Employment Effects of Continuous Off-the-job Training in East Germany after Unification, *Journal of Business Economics and Statistics*, 17, 74–90.

Lechner, M. (2001): Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption, in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.

Lechner, M. (2002): Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programs by Matching Methods, *Journal of the Royal Statistical Society, Series A*, 165, 59–82.

Lechner, M. (2018): Modified Causal Forests for Estimating Heterogeneous Causal Effects. Version 2. arXiv: 1812.09487v2.

Lechner, M. and C. Wunsch (2013): Sensitivity of matching based program evaluations to the availability of control variables, *Labour Economics* 21, 111-121.

Lechner, M., R. Miquel, and C. Wunsch (2011): Long-run effects of public sector sponsored training in West-Germany, *Journal of the European Economic Association*, 9 (4), 742–784.

Rosenbaum, P.R., and D.B. Rubin (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, 70, 41-50.

Rosenbaum, P.R., and D.B. Rubin (1985): Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score, *The American Statistician*, 39, 33-38.

Rubin, D. B. (1974): Estimating Causal Effects of Treatments in Randomized and Nonrandomized *Studies, Journal of Educational Psychology, 66, 688-701.*

Sianesi, B. (2004): An evaluation of the Swedish system of active labour market programs in the 1990s, *Review of Economics and Statistics* 86, 133–155.

Sianesi, B. (2008): Differential effects of active labour market programs for the unemployed, *Labour Economics*, 15 (3), 370–399.

Van den Berg, G., and J. Vikström (2019): Long-Run Effects of Dynamically Assigned Treatments: A New Methodology and an Evaluation of Training Effects on Earnings, IZA Discussion paper 12470, Bonn: IZA.

Van Hooft, E. A. J., and G. Noordzij (2009): The Effects of Goal Orientation on Job Search and Reemployment: A Field Experiment Among Unemployed Job Seekers, *Journal of Applied Psychology* 94(6), 1581-1590.

Vikström, J. (2017): Dynamic treatment assignment and evaluation of active labor market policies. *Labour Economics*, 49, 42–54.

Wager, S., and S. Athey (2018): Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, *Journal of the American Statistical Association*, 113:523, 1228-1242.

Zhou, Z., S. Athey, and S. Wager (2018): Offline Multi-Action Policy Learning: Generalization and Optimization, arXiv:1810.04778v2.

Appendix A: Data

Table A.1: Means and standardized differences for conditioning variables and outcomes

	No ALMP participation (NOP)[3]	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
A. Conditioning variables				
(Pseudo-) Duration until first treatment, in days (Daction)	92.66	124.39	122.81	119.17
		(51.73)	(49.29)	(43.62)
Gender (female =1)(Woman)	.49	.31	.40	.47
		(35.73)	(16.48)	(3.16)
Age in years (age)	35.09	34.01	34,00	33.63
		(11.87)	(12.28)	(16.53)
Living in a city (city)	.36	.37	.32	.31
		(2.43)	(6.98)	(11.02)
Knowledge of french (frans)	.63	.54	.72	.67
		(17.87)	(20.85)	(9.07)
Knowledge of english (engels)	.70	.66	.82	.79
		(7.86)	(29.32)	(20.30)
Knowledge of german (duits)	.25	.19	.34	.32
		(13.20)	(19.35)	(16.31)
Knowledge of italian (italiaans)	.03	.02	.02	.03
		(6.85)	(4.25)	(3.92)
Knowledge of spanish (spaans)	.08	.05	.07	.07
		(12.83)	(2.42)	(4.02)
Proficiency in dutch ² (Lang_dutch)	2.44	2.51	2.72	2.64
		(8.68)	(39.74)	(27.10)
Ever been in BIT before current unemployment spell (cat_2)	.33	.36	.43	.40
		(5.04)	(20.62)	(14.14)
Ever been in unemployment without U-benifit before current unemployment spell (cat_3)	.24	.29	.24	.25
		(9.64)	(1.53)	(2.39)
Ever been on welfare before current unemployment spell (cat_5)	.07	.14	.04	.08
		(22.27)	(11.62)	(4.27)
Having had an unemployment benifit sanction before current unemployment spell (cat_14)	.06	.07	.04	.03
		(6.63)	(9.15)	(12.71)
Ever have had a sickness benifit before current unemployment spell (cat_76)	.12	.14	.10	.11
		(4.52)	(8.43)	(4.86)
Ever been back to education before current unemployment spell (cat 77)	.04	.04	.06	.06
		(1.08)	(10.59)	(9.76)

Table A.1, continued

Ever been part time working, part time unemployed before current unemployment spell (cat_80)	.19	.15 (9.68)	.14 (12.74)	.17 (3.20)
Ever been in BIT and part time work before current unemployment spell (cat_82)	.05	.03 (11.43)	.05 (2.37)	.06 (4.67)
Ever been in on-the-job-training before current unemployment spell (cat_85)	.11	.16 (14.76)	.16 (16.18)	.17 (18.82)
Ever been in temporary agency work before current unemployment spell (cat_89)	.21	.28 (16.80)	.27 (14.02)	.24 (7.47)
Ever been working full-time but looking for another job before current unemployment spell (cat_90)	.29	.33 (8.76)	.42 (27.23)	.40 (23.20)
Ever been working part time + part time in education before current unemployment spell (cat_91)	.01	.02 (8.68)	.01 (7.36)	.02 (2.10)
Ever been working part time and part time looking for a job before current unemployment spell (cat_93)	.18	.14 (11.86)	.15 (6.84)	.18 (1.33)
Ever have had limited search obligations because of family or social reasons before current unemployment spell (cat_96)	.02	.01 (.86)	.01 (7.73)	.02 (.29)
Ever have had limited search obligations because participation in training before current unemployment spell (cat_97)	.03	.03 (.84)	.04 (4.67)	.04 (3.08)
Number of months in unemployment in the 10 years before current spell (Unem_10jaar)	18.05	19.12 (5.44)	15.91 (11.14)	17.3 (3.80)
Number of months with sickness benefit in the 10 years before current spell (ziek_10jaar)	1.05	.99 (1.09)	.72 (7.27)	1.03 (.32)
Number of months unknown position in the 10 years before current spell (mystery_10jaar)	9.11	9.44 (1.53)	9.91 (3.55)	10.17 (4.76)
Number of months of work in the 10 years before current spell (werk_10jaar)	48.69	54.18 (14.35)	59.48 (28.13)	54.89 (16.24)
Number of months in unemployment in the 5 years before current spell (unem_5jaar)	10.52	11.24 (5.80)	8.91 (13.63)	9.81 (5.88)
Number of months with sickness benefit in the 5 years before current spell (ziek_5jaar)	.66	.50 (5.11)	.48 (5.94)	.60 (1.79)
Number of months in unknown position in the 5 years before current spell (mystery_5jaar)	2.89	3.32 (4.49)	3,00 (1.15)	3.39 (5.01)

Table A.1, continued

Number of months of work in the 5 years before current spell (werk_5jaar)	31.81	35.30 (16.87)	38.64 (33.31)	36.3 (21.77)
Number of months in unemployment in the 2 years before current spell (unem_2jaar)	3.89	3.77 (2.18)	2.97 (18.01)	3.18 (13.79)
Number of months with sickness benefit in the 2 years before current spell (ziek_2jaar)	.22	.16 (4.39)	.19 (2.02)	.17 (3.24)
Number of months in unknown position in the 2 years before current spell (mystery_2jaar)	.53	.64 (3.71)	.58 (1.50)	.67 (4.58)
Number of months of work in the 2 years before current spell (werk_2jaar)	15.50	17.38 (22.57)	18.50 (36.82)	18.00 (30.32)
Having experience in the preferred profession [4] (exp)	1.63	1.86 (21.21)	1.73 (8.73)	1.80 (14.90)
Duration previous job in months (duur_laatste_werk)	17.10	19.86 (20.66)	21.50 (32.67)	20.69 (26.87)
Number of professions the person is interested in (aant_beroepen)	2.77	3.04 (14.35)	2.98 (10.86)	3.01 (12.65)
Having participated in a training before current spell (vroeger_o)	.08	.12 (13.02)	.11 (10.77)	.11 (10.71)
Having participated in a training course dutch before current spell (vroeger_n)	.01	.03 (10.78)	.01 (3.38)	.01 (.05)
Having participated in a on-the-job-training before current spell (vroeger_i)	.07	.10 (11.25)	.11 (13.29)	.12 (17.30)
Having participated in intensive couselling before current spell (vroeger_t)	.06	.08 (5.04)	.05 (4.38)	.06 (.31)
Having participated in a orientation training before current spell (vroeger_r)	.01	.01 (.09)	.03 (9.61)	.03 (10.54)
Having participated in another ALMP before current spell (vroeger_a)	.04	.07 (13.53)	.05 (4.37)	.04 (3.04)
Drivers license car (rijbew_B)	.68	.63 (10.79)	.76 (16.39)	.71 (5.91)
Drivers license truck (rijbew_C)	.04	.05 (4.81)	.03 (4.98)	.03 (6.19)
Divers license bus (rijbew_D)	.01	.01 (2.84)	.01 (.98)	.01 (2.46)
Educ. Attainment: first level secondary education or lower (so1)	.14	.18 (10.29)	.04 (34.75)	.10 (13.63)

Table A.1, continued

Educ. Attainment: part time education, professional track (dbso)	.04	.06 (9.51)	.03 (2.97)	.04 (1.73)
Educ. Attainment; 2nd level secondary education, general & artistic tracks (akso2)	.02	.02 (1.72)	.02 (.84)	.03 (4.66)
Educ. Attainment; 2nd level secondary education, professional track (bso2)	.09	.13 (13.53)	.06 (9.47)	.08 (2.65)
Educ. Attainment: secundary education for pupils with special needs (buso)	.02	.04 (11.29)	.01 (5.18)	.02 (1.69)
Educ. Attainment; 2nd level secondary education, technological track (tso2)	.04	.05 (4.11)	.04 (.02)	.05 (8.08)
Educ. Attainment; 3rd level secondary education, general & artistic tracks (akso3)	.08	.08 (1.39)	.13 (16.86)	.11 (10.42)
Educ. Attainment; 3rd level secondary education, professional track (bso3)	.21	.26 (13.3)	.21 (1.41)	.22 (4.08)
Educ. Attainment; 3rd level secondary education, technological track (tso3)	.11	.11 (.27)	.22 (29.68)	.17 (17.82)
Educ. Attainment: higher professional education (hbo)	.02	.01 (8.48)	.02 (4.01)	.02 (.52)
Educ. Attainment: professional bachelor (pba)	.13	.04 (31.38)	.14 (4.19)	.11 (4.89)
Educ. Attainment: academic bachelor (aba)	.02	0.00 (11.01)	.01 (2.08)	.01 (1.02)
Educ. Attainment: master (ma)	.10	.02 (32.94)	.06 (16.26)	.04 (24.03)
District of residence: Antwerpen (Antwerpen)	.23	.18 (11.78)	.17 (15.21)	.11 (31.53)
District of residence: Mechelen (Mechelen)	.05	.04 (5.04)	.04 (1.23)	.04 (2.37)
District of residence: Turnhout (Turnhout)	.06	.08 (6.27)	.09 (9.15)	.03 (13.70)
District of residence: Leuven (Leuven)	.06	.08 (4.83)	.07 (3.41)	.05 (4.95)
District of residence: Vilvoorde (Vilvoorde)	.09	.04 (16.56)	.04 (17.03)	.05 (14.31)
District of residence: Brugge (Brugge)	.04	.07 (13.63)	.05 (4.33)	.05 (4.78)
District of residence: Ieper or Diksmuide (Ieper)	.02	.03 (7.74)	.03 (6.16)	.03 (3.97)
District of residence: Kortrijk (Kortrijk)	.03	.04 (2.85)	.04 (3.76)	.05 (8.18)

Table A.1, continued

District of residence: Oostende (Oostende)	.02	.06	.04	.05
		(16.33)	(7.81)	(13.80)
District of residence: Roeselare (Roeselare)	.02	.02	.03	.03
		(2.97)	(7.13)	(7.55)
District of residence: Tielt or Veurne (Tielt)	.02	.02	.02	.02
		(5.34)	(3.06)	(2.45)
District of residence: Aalst (Aalst)	.04	.04	.03	.04
		(1.27)	(9.14)	(.88)
District of residence: Dendermonde (Dendermonde)	.03	.03	.03	.03
		(3.77)	(.54)	(1.18)
District of residence: Eeklo (Eeklo)	.01	.01	.01	.01
		(1.68)	(4.94)	(1.27)
District of residence: Gent (Gent)	.09	.06	.09	.09
		(10.95)	(.87)	(1.06)
District of residence: Oudenaarde (Oudenaarde)	.02	.01	.00	.03
		(9.38)	(11.69)	(6.14)
District of residence: Sint-Niklaas (Sintniklaas)	.04	.03	.04	.02
		(3.83)	(.50)	(10.91)
District of residence: Hasselt (Hasselt)	.07	.09	.11	.11
		(5.16)	(14.26)	(13.66)
District of residence: Maaseik (Maaseik)	.04	.05	.05	.08
		(4.27)	(7.98)	(17.49)
District of residence: Tongeren (Tongeren)	.04	.03	.03	.09
		(2.39)	(1.59)	(22.58)
Country of birth: Belgium (belg)	.65	.66	.77	.71
		(2.09)	(28.05)	(13.01)
Country of birth: Western & Northern EU (eu_core)	.07	.05	.06	.06
		(5.45)	(1.42)	(1.16)
Country of birth: Southern EU (eu_south)	.01	.01	.01	.02
		(3.04)	(6.33)	(3.00)
Country of birth: Eastern EU (eu_east)	.06	.04	.03	.03
		(10.12)	(13.22)	(12.13)
Country of birth: Turkey & Morocco (tm)	.05	.04	.02	.03
		(8.27)	(16.19)	(9.71)
Country of birth: Rest of the world (row)	.16	.21	.10	.14
		(11.68)	(17.18)	(4.68)
Calendar month start unemployment spell 1 (StartUI1)	.05	.04	.04	.06
		(7.93)	(7.04)	(3.39)
Calendar month start unemployment spell 2 (StartUI2)	.08	.06	.09	.08
		(5.41)	(2.83)	(1.55)
Calendar month start unemployment spell 3 (StartUI3)	.05	.06	.04	.07
		(5.20)	(.82)	(9.22)

Table A.1, continued

Calendar month start unemployment spell 4	.05	.05	.06	.06
(StartUI4)		(.82)	(4.67)	(5.12)
Calendar month start unemployment spell 5	.05	.05	.06	.04
(StartUI5)		(.02)	(4.79)	(3.47)
Calendar month start unemployment spell 6	.04	.05	.04	.04
(StartUI6)		(4.70)	(.93)	(2.01)
Calendar month start unemployment spell 7	.05	.06	.06	.05
(StartUI7)		(4.15)	(5.87)	(1.17)
Calendar month start unemployment spell 8	.07	.06	.07	.06
(StartUI8)		(2.77)	(.66)	(2.38)
Calendar month start unemployment spell 9	.06	.05	.06	.05
(StartUI9)		(4.34)	(1.1)	(3.39)
Calendar month start unemployment spell 10	.07	.07	.07	.05
(StartUI10)		(1.17)	(2.49)	(7.78)
Calendar month start unemployment spell 11	.06	.07	.06	.06
(StartUI11)		(3.20)	(1.67)	(.42)
Calendar month start unemployment spell 12	.05	.05	.04	.05
(StartUI12)		(2.67)	(7.03)	(3.44)
Calendar month start unemployment spell 13	.05	.06	.04	.05
(StartUI13)		(4.77)	(4.76)	(2.23)
Calendar month start unemployment spell 14	.07	.07	.05	.05
(StartUI14)		(1.22)	(7.33)	(5.60)
Calendar month start unemployment spell 15	.05	.06	.04	.05
(StartUI15)		(5.45)	(2.35)	(.83)
Calendar month start unemployment spell 16)	.04	.04	.04	.05
(StartUI16		(.64)	(.47)	(3.98)
Calendar month start unemployment spell 17	.04	.03	.05	.04
(StartUI17)		(2.49)	(6.41)	(2.38)
Calendar month start unemployment spell 18	.04	.03	.05	.04
(StartUI18)		(2.16)	(4.51)	(1.79)
Calendar month start unemployment spell 19	.04	.04	.05	.03
(StartUI19)		(.09)	(3.08)	(4.86)
Preferred profession 2 [5] (prof2)	.10	.04	.20	.20
		(25.87)	(29.63)	(29.52)
Preferred profession 19 [5] (prof19)	.05	.09	.05	.05
		(16.09)	(.39)	(2.48)

Table A.1, continued

Preferred profession 21 [5] (prof21)	.03	0,00	.02	.01
		(20.17)	(5.93)	(14.05)
Preferred profession 24 [5] (prof24)	.02	.01	.02	.02
		(10.78)	(5.95)	(6.21)
Preferred profession 25 [5] (prof25)	.05	.06	.05	.06
		(3.85)	(2.10)	(5.01)
Preferred profession 26 [5] (prof26)	.04	.05	.04	.04
		(3.21)	(3.93)	(3.08)
Preferred profession 27 [5] (prof27)	.02	.01	.01	.01
		(3.29)	(3.87)	(9.65)
Preferred profession 29 [5] (prof29)	.05	.07	.03	.03
		(7.66)	(13.38)	(11.40)
Preferred profession 30 [5] (prof30)	.01	.02	.01	.01
		(3.46)	(.25)	(2.07)
Preferred profession 31 [5] (prof31)	.07	.12	.06	.06
		(18.21)	(.68)	(3.04)
Preferred profession 32 [5] (prof32)	.04	.01	.06	.06
		(15.57)	(12.47)	(11.31)
Preferred profession 33 [5] (prof33)	.02	.01	.04	.04
		(10.63)	(10.59)	(8.46)
Preferred profession 34 [5] (prof34)	.02	.02	.01	.01
		(1.41)	(7.56)	(3.33)
Preferred profession 35 [5] (prof35)	.01	.02	.01	.02
		(8.81)	(2.30)	(7.52)
Preferred profession 36 [5] (prof36)	.01	.01	.01	.01
		(4.18)	(2.72)	(2.55)
Preferred profession 37 [5] (prof37)	.02	.03	.03	.03
		(5.92)	(5.33)	(5.98)
Preferred profession 38 [5] (prof38)	.01	.01	.03	.02
		(.95)	(10.76)	(5.56)
Preferred profession 39 [5] (prof39)	.01	.01	.02	.01
		(.63)	(7.66)	(1.66)
Preferred profession 40 [5] (prof40)	.05	.05	.03	.03
		(2.33)	(12.74)	(8.74)
Preferred profession 41 [5] (prof41)	.02	0,00	.02	.02
		(16.53)	(6.20)	(.01)
Preferred profession 42 [5] (prof42)	.18	.26	.18	.18
		(19.08)	(.44)	(2.02)
Preferred profession 43 [5] (prof43)	.07	.08	.03	.03
		(4.95)	(20.42)	(16.08)
Preferred profession 44 [5] (prof44)	.09	.02	.05	.03
		(32.7)	(15.81)	(23.52)
Economic sector of previous job 0 [6] (sect0)	.30	.31	.27	.29
		(.57)	(7.40)	(2.74)

Table A.1, continued

Economic sector of previous job 1 [6] (sect1)	.01	.02	.01	.01
		(7.79)	(1.12)	(.65)
Economic sector of previous job 2 [6] (sect2)	.01	.01	.01	.02
		(3.74)	(.15)	(.98)
Economic sector of previous job 3 [6] (sect3)	.03	.03	.02	.03
		(1.09)	(4.79)	(1.52)
Economic sector of previous job 4 [6] (sect4)	.01	.02	.01	.01
		(6.09)	(2.33)	(3.01)
Economic sector of previous job 5 [6] (sect5)	.03	.03	.02	.04
		(1.31)	(9.20)	(4.84)
Economic sector of previous job 6 [6] (sect6)	.04	.01	.02	.01
		(18.70)	(12.32)	(17.36)
Economic sector of previous job 7 [6] (sect7)	.01	.02	.01	.02
		(3.28)	(3.47)	(1.41)
Economic sector of previous job 8 [6] (sect8)	.02	.02	.02	.02
		(1.27)	(2.02)	(3.01)
Economic sector of previous job 9 [6] (sect9)	.01	.01	.01	.00
		(2.03)	(5.20)	(9.56)
Economic sector of previous job 10 [6] (sect10)	.01	.01	.03	.02
		(1.11)	(8.62)	(2.18)
Economic sector of previous job 11 [6] (sect11)	.02	.03	.02	.02
		(7.94)	(.79)	(4.54)
Economic sector of previous job 12 [6] (sect12)	.01	.03	.02	.01
		(11.83)	(4.21)	(2.80)
Economic sector of previous job 13 [6] (sect13)	.01	.01	.02	.01
		(3.40)	(1.62)	(.16)
Economic sector of previous job 14 [6] (sect14)	.01	.01	.01	.02
		(.51)	(.55)	(6.46)
Economic sector of previous job 15 [6] (sect15)	.02	.02	.02	.02
		(1.01)	(1.41)	(1.13)
Economic sector of previous job 16 [6] (sect16)	.01	.01	.01	.01
		(5.55)	(2.44)	(5.10)
Economic sector of previous job 17 [6] (sect17)	.01	0,00	0,00	0,00
		(6.72)	(6.28)	(6.87)
Economic sector of previous job 18 [6] (sect18)	.02	.02	.03	.03
		(1.33)	(2.92)	(5.99)
Economic sector of previous job 19 [6] (sect19)	.01	.01	.02	.02
		(.92)	(8.82)	(3.83)
Economic sector of previous job 20 [6] (sect20)	.01	.02	.02	.01
		(1.79)	(6.13)	(4.33)
Economic sector of previous job 21 [6] (sect21)	.03	.04	.04	.03
		(4.90)	(5.11)	(1.08)
Economic sector of previous job 22 [6] (sect22)	.02	.02	.02	.03
		(2.42)	(1.91)	(3.90)

Table A.1, continued

Economic sector of previous job 23 [6] (sect23)	.02	.03	.01	.01
		(6.04)	(3.87)	(5.17)
Economic sector of previous job 24 [6] (sect24)	.02	.02	.03	.02
		(5.74)	(8.77)	(3.64)
Economic sector of previous job 25 [6] (sect25)	.01	.01	.01	.01
		(2.72)	(6.91)	(1.75)
Economic sector of previous job 26 [6] (sect26)	.01	.01	.01	.01
		(1.82)	(6.46)	(4.73)
Economic sector of previous job 27 [6] (sect27)	.02	.01	.01	.01
		(6.63)	(9.14)	(3.42)
Economic sector of previous job 28 [6] (sect28)	.01	.02	.02	.02
		(6.81)	(8.84)	(5.50)
Economic sector of previous job 29 [6] (sect29)	.02	.04	.04	.03
		(10.86)	(10.77)	(6.08)
Economic sector of previous job 30 [6] (sect30)	.01	.02	.01	.01
		(4.86)	(2.39)	(1.08)
Economic sector of previous job 31 [6] (sect31)	.01	.02	.01	.01
		(7.53)	(.86)	(.92)
Economic sector of previous job 32 [6] (sect32)	.02	.02	.01	.01
		(1.29)	(3.36)	(7.06)
Economic sector of previous job 33 [6] (sect33)	.03	.01	.05	.04
		(11.67)	(7.34)	(5.03)
Economic sector of previous job 34 [6] (sect34)	.04	.02	.05	.04
		(12.25)	(2.30)	(.82)
Economic sector of previous job 35 [6] (sect35)	.06	.04	.06	.07
		(9.90)	(1.35)	(5.45)
Economic sector of previous job 36 [6] (sect36)	.02	.02	.03	.02
		(5.95)	(2.73)	(5.31)
Number of kids=0 (Nkids0)	.12	.17	.14	.17
		(16.13)	(7.96)	(16.31)
Number of kids=1 (Nkids1)	.04	.04	.04	.04
		(3.24)	(3.22)	(3.32)
Number of kids=2 (Nkids2)	.03	.03	.03	.03
		(1.51)	(2.39)	(.91)
Number of kids > 3 (Nkids3)	.02	.01	.01	.01
		(4.71)	(7.42)	(2.27)
Number of kids missing (Nkids4)	.80	.75	.79	.75
		(10.19)	(2.15)	(11.15)
Household head: no (head0)	.45	.48	.49	.48
		(5.31)	(7.75)	(5.00)
Household head: yes (head1)	.09	.08	.06	.07
		(3.58)	(11.06)	(6.95)
Household head: unknown (head2)	.45	.44	.44	.45
		(3.28)	(1.79)	(1.14)

Table A.1, continued

B. Outcomes				
Cumulative months in employment 9 months after programme start (Cwa9)	3.51	3.35	2.34	2.03
		(4.83)	(36.91)	(46.35)
Cumulative months in unemployment 9 months after programme start (Cua9)	4.90	5.44	6.44	6.68
		(16.16)	(49.18)	(56.21)
Cumulative months out of the labour force 9 months after programme start (Cia9)	.58	.21	.22	.29
		(25.63)	(25.93)	(20.23)
Cumulative months in employment 30 months after programme start (Cwa30)	16.04	18.45	17.11	14.82
		(24.06)	(10.83)	(11.92)
Cumulative months in unemployment 30 months after programme start (Cua30)	10.83	10.11	11.75	13.47
		(8.42)	(10.66)	(29.89)
Cumulative months out of the labour force 30 months after programme start (Cia30)	3.13	1.44	1.14	1.71
		(27.71)	(34.03)	(23.14)
Number of observations	59,964	1,305	1,22	1,115

Notes: [1] The standardized difference is $\left|\bar{x}^{j} - \bar{x}^{NOP}\right| / \sqrt{[Var(x^{j}) + Var(x^{NOP})]/2}$, where $\bar{x}^{j}$ and $Var(x^{j})$ are the sample mean and variance of the variable x^{j} for $j \in \{SVT, LVT, DLT, IC, OT\}$.

[2] Proficiency in Dutch = 0 if no knowledge; = 1 if limited; = 2 if good; =3 if very good.

[3] For non-participants in ALMP (NOP) the date at which the ALMP starts is predicted (See Section 4.4).

[4] Experience: no experience (0), limited experience (1), good experience (2), a lot of experience (3)

[5] 2 = General clerk; 19 = Goods handlers; 21 = Managers of a department or service; 24 = Educators; 25 = Sales support staff; 26 = Sales representatives; 27 = Representatives; 29 = Hall staff; 30 = Food workers; 31 = Construction workers and technicians; 32 = Specialised administrative staff; 33 = Computer and ICT staff; 34 = Agricultural, horticultural and forestry workers and fishermen; 35 = Vehicle mechanics; 36 = Staff involved in tourism, leisure and sport; 37 = Metalworkers; 38 = Draughtsmen and designers; 39 = Transport and logistics personnel; 40 = Nurses and carers, Law enforcement and rescue workers, Medics, Paramedics and laboratory assistants; 41 = Private sector consultants, Bank and insurance experts, Business consultants; 42 = Craftsmen, Drivers, Apparel and leatherworkers, Miscellaneous production workers, Printers, Electricians and electricians; Woodworkers, Industrial technicians, Machinists and crane operators; Metal production workers, Operators chemistry and plastics, Precision technicians, Technical managers, Textile workers, Rail, water and air transport workers; 43 = Personal service providers, Cleaning and maintenance personnel; 44 = Architects and surveyors, Artists, artists and other cultural professions, Controllers and inspectors, Instruction, training and education personnel, Media personnel, Teaching and management personnel in schools, Researchers and experts study service, Socio-cultural workers; [6] 0 = Missing; 1 = Construction of houses; 2 = Retail sale of clothing; 3 = Dining facility full service; 4 = Dining facility limited service; 5 = General cleaning of buildings; 6 = Ordinary general secondary education; 7 = Other social work; 8 = Retail sale of food; 9 = Institutions for the elderly and disabled; 10 = Electrical installation, plumbing and other construction installation; 11 = Finishing of buildings; 12 = Other specialised construction activities; 13 = Wholesale trade in consumer goods; 14 = Retail trade in consumer goods; 15 = Retail trade in other goods; 16 = Other social work activities without accommodation; 17 = Other personal services; 18 = Manufacture of food products; 19 = Manufacture of fabricated metal products, except machinery and equipment; 20 = Wholesale and retail trade, maintenance and repair of motor vehicles and motorcycles; 21 = Wholesale trade and commission trade, except of motor vehicles and motorcycles; 22 = Retail trade, except of motor vehicles and motorcycles; 23 = Land transport and transport via pipelines; 24 = Warehousing and support activities for transportation; 25 = Food and beverage service activities; 26 = Services to buildings and landscape gardening; 27 = Education; 28 = Industry 1; 29 = Industry 2; 30 = Industry 3, Electricity, gas, steam and air conditioning supply, Water supply; Waste and sewerage management and remediation services; 31 = Construction, wholesale and retail trade; Repair of motor vehicles and motorcycles; 32 = Transport and storage, Accommodation and food service activities;33 = Information and communication, Financial and insurance activities, Real estate activities, Professional, scientific and technical activities;34 = Administrative and support service activities 1; 35 = Administrative and support service activities 2, Public administration and defence; compulsory social security, Human health and social work activities; 36 = Arts, entertainment and recreation, Other services.

Appendix B: Part of econometrics

B.1 Predictions of the pseudo durations for the NOP

For all participants in training log duration until the programme start is regressed on all the explanatory variables, including the interactions of all the explanatory variables with gender. This was done with a LASSO regression, using the R-package glmnet (Friedman et al., 2008). A ten-fold cross validation approach is used to determine the penalty term. As the LASSO estimates are biased, the regression is re-run using the subset of variables selected by the LASSO, to obtain OLS-estimates (Post-Lasso). Subsequently, the OLS-results are used to predict the duration until start in the NOP sample. Then we add to this prediction a draw from a Normal distribution with mean zero and the standard error being the one of the OLS regression.

The initial NOP sample has 112,128 observations. For 54,255 observations, the duration of the unemployment spell is smaller than the predicted duration until start. For 6,732 observations, the predicted duration until start exceeds nine months. For 5,371 observations, both problems are present. As a result, there are 55,616 observations where either or both problems apply. These observations are dropped from the sample.

The LASSO regression is not reported while the Post LASSO regression is reported in Table B.1.

Table B.1: Post LASSO regression of the log duration until the programme start on the explanatory variables in a LASSO regression

Intercept	4,96	(0,16)	***
Number of months in unemployment in the 10 years before current spell (Unem_10jaar)	0,00	(0,00)	***
Number of months unknown position in the 10 years before current spell (mystery_10jaar)	0,00	(0,00)	*
Number of months out of the labour force in the 10 years before current spell (olf_10jaar)	0,00	(0,00)	
Number of months of work in the 5 years before current spell (werk_5jaar)	0,00	(0,00)	*
Number of months in unemployment in the 2 years before current spell (unem_2jaar)	-0,01	(0,00)	*
Number of months unknown position in the 2 years before current spell (mystery_2jaar)	0,00	(0,00)	
Number of months out of the labour force in the 2 years before current spell (olf_2jaar)	0,00	(0,01)	
Drivers license car (rijbew_B)	-0,02	(0,03)	
Drivers license truck (rijbew_C)	-0,19	(0,06)	***
Educ. Attainment: part time education, professional track (dbso)	0,03	(0,05)	
Educ. Attainment; 3rd level secondary education, general track (aso3)	0,03	(0,03)	
Educ. Attainment; 3rd level secondary education, artistic track (kso3)	0,08	(0,08)	
Educ. Attainment: master (ma)	0,05	(0,07)	
Age in years (age)	-0,01	(0,03)	
Knowledge of english (engels)	-0,03	(0,02)	
Knowledge of german (duits)	0,02	(0,03)	
Proficiency in dutch very good (ned3)	0,06	(0,02)	**
District of residence: Antwerpen (Antwerpen)	-0,13	(0,08)	*
District of residence Mechelen(Mechelen)	0,08	(0,05)	
District of residence Turnhout (Turnhout)	0,07	(0,05)	
District of residence Leuven (Leuven)	0,13	(0,06)	**
District of residence Vilvoorde (Vilvoorde)	0,14	(0,05)	***
District of residence Brugge (Brugge)	-0,22	(0,07)	***
District of residence Ieper (Ieper)	-0,15	(0,07)	**
District of residence Kortrijk (Kortrijk)	0,06	(0,06)	
District of residence Dendermonde (Dendermonde)	-0,13	(0,08)	
District of residence Gent (Gent)	0,14	(0,05)	***
District of residence Oudenaarde (Oudenaarde)	-0,04	(0,11)	
District of residence Hasselt (Hasselt)	-0,03	(0,05)	
District of residence Maaseik (Maaseik)	0,09	(0,07)	
District of residence Tongeren (Tongeren)	-0,03	(0,06)	

Table B.1, continued

Household-head	-0,03	(0,01)	**
Number of professions the person is interested in (aant_beroepen)	0,00	(0,01)	
Ever been in BIT before current unemployment spell (cat_2)	-0,02	(0,03)	
Ever been in unemployment without U-benifit before current unemployment spell (cat_3)	-0,01	(0,02)	
Ever been in PWA before current unemployment spell (cat_30)	0,15	(0,25)	
Ever been in a trajectory from sickness benefit to work before current unemployment spell (cat_32)	-0,15	(0,10)	
Ever been in Arbeidszorg before current unemployment spell (cat_33)	-1,28	(0,47)	***
Ever been in BIT and part time work before current unemployment spell (cat_82)	-0,02	(0,04)	
Ever been in temporary agency work before current unemployment spell (cat_89)	-0,11	(0,03)	***
Ever been working full-time but looking for another job before current unemployment spell (cat_90)	-0,08	(0,02)	***
Ever been working part time + part time in education before current unemployment spell (cat_91)	-0,14	(0,10)	
Ever been in the specific status given to some high skilled unemployed from outside the European Economic Area, before current unemployment spell (cat_94)	0,10	(0,16)	
Ever have had limited search obligations because of family or social reasons before current unemployment spell (cat_96)	0,06	(0,36)	
Ever have had limited search obligations because participation in training before current unemployment spell (cat_97)	-0,04	(0,05)	
Unemployment rate in district of residence (Wlgr)	-2,42	(1,59)	
Educational attainment high (High)	0,02	(0,05)	
Age between >=22, < 25 (age_lt25)	-0,05	(0,03)	
Age between >=36, < 50 (age_lt50)	0,06	(0,05)	
Age between >=50 and <=55 (age_ge50)	0,13	(0,08)	
Country of birth: Belgium (belg)	-0,03	(0,03)	
Country of birth: Western & Northern EU (eu_core)	0,04	(0,06)	
Unemployment spell began in January (m1)	-0,12	(0,06)	**
Unemployment spell began in February (m2)	-0,22	(0,06)	***
Unemployment spell began in March (m3)	0,02	(0,03)	
Unemployment spell began in June (m6)	0,06	(0,05)	
Unemployment spell began in July (m7)	-0,05	(0,06)	
Unemployment spell began in August (m8)	-0,36	(0,06)	***
Unemployment spell began in September (m9)	-0,27	(0,05)	***
Unemployment spell began in October (m10)	-0,08	(0,05)	
Unemployment spell began in November (m11)	-0,16	(0,04)	***
Unemployment spell began in December (m12)	-0,16	(0,04)	***
Duration previous job in months (duur_laatste_werk)	0,00	(0,00)	**
Having participated in a on-the-job-training before current spell (vroeger_i)	-0,02	(0,03)	

Table B.1, continued

Having participated in a dutch training course before current spell (vroeger_n)	-0,02	(0,08)	
Having participated in a training before current spell (vroeger_o)	-0,04	(0,04)	
Having participated in an orientation training before current spell (vroeger_r)	-0,10	(0,07)	
Having participated in intensive couselling before current spell (vroeger_t)	0,03	(0,04)	
Having limited experience in the preferred profession (ervaring_beperkt)	-0,02	(0,03)	
Having no experience in the preferred profession (ervaring_geen)	0,00	(0,04)	
Knowledge of another language (different from Dutch, English, German, Italian or Spanish) (andere_taal)	0,01	(0,04)	
District or residence Aalst, interaction with sex	-0,05	(0,07)	
District or residence Antwerpen, interaction with sex	0,13	(0,06)	**
District or residence Brugge, interaction with sex	0,22	(0,09)	**
District or residence Dendermonde, interaction with sex	0,17	(0,11)	
District or residence Eeklo, interaction with sex	-0,10	(0,11)	
District or residence Gent, interaction with sex	-0,08	(0,06)	
District or residence Leuven, interaction with sex	-0,20	(0,08)	***
District or residence Maaseik, interaction with sex	-0,13	(0,09)	
District or residence Oostende, interaction with sex	0,06	(0,10)	
District or residence Oudenaarde, interaction with sex	-0,27	(0,15)	*
District or residence Roeselare, interaction with sex	-0,10	(0,10)	
District or residence Sint-Niklaas, interaction with sex	0,08	(0,08)	
District or residence Turnhout, interaction with sex	0,07	(0,07)	
Educ. Attainment: academic bachelor, interaction with sex	-0,13	(0,15)	
Age between >=50 and <=55, interaction with sex	0,05	(0,08)	
Age between >=36, < 50, interaction with sex	-0,10	(0,04)	**
Knowledge of another language (different from Dutch, English, German, Italian or Spanish, interaction with sex	0,01	(0,05)	
Educ. Attainment; 2nd level secondary education, general track, interaction with sex	-0,17	(0,12)	
Educ. Attainment; 3rd level secondary education, professional track , interaction with sex	-0,03	(0,04)	
Having had an unemployment benifit sanction before current unemployment spell, interaction with sex	0,00	(0,07)	
Ever have had a sickness benifit before current unemployment spell, interaction with sex	-0,09	(0,04)	**
Ever been back to education before current unemployment spell, interaction with sex	-0,01	(0,06)	
Ever been part time working, part time unemployed before current unemployment spell, interaction with sex	0,03	(0,03)	
Ever been in on-the-job-training before current unemployment spell, interaction with sex	-0,03	(0,05)	

Table B.1, continued

Ever been in temporary agency work before current unemployment spell, interaction with sex	0,10	(0,04)	**
Ever been working part time + part time in education before current unemployment spell, interaction with sex	0,14	(0,15)	
Ever been in temporary unemployment before current unemployment spell, interaction with sex	0,28	(0,24)	
Ever been working part time and part time looking for a job before current unemployment spell, interaction with sex	-0,06	(0,03)	**
Ever have had limited search obligations because of family or social reasons before current unemployment spell, interaction with sex	0,08	(0,37)	
Living in a city, interaction with sex	0,03	(0,03)	
Educ. Attainment: part time education, professional track, interaction with sex	0,05	(0,09)	
Knowledge of German, interaction with sex	0,09	(0,05)	**
No experience in preferred profession, interaction with sex	0,13	(0,05)	**
Good experience in preferred profession, interaction with sex	0,04	(0,03)	
Country of birth: Western & Northern EU, interaction with sex	0,06	(0,08)	
Country of birth: Southern EU, interaction with sex	0,18	(0,13)	
Knowledge of French, interaction with sex	0,07	(0,03)	**
Educ. Attainment: higher professional education , interaction with sex	-0,10	(0,12)	
Knowledge of Italian , interaction with sex	-0,12	(0,09)	
Educational attainment, secondary, 3rd level, interaction with sex	-0,03	(0,06)	
Educational attainment is low , interaction with sex	0,02	(0,06)	
Unemployment spell began in January, interaction with sex	-0,29	(0,08)	***
Unemployment spell began in February, interaction with sex	0,19	(0,09)	**
Unemployment spell began in April, interaction with sex	0,01	(0,05)	
Unemployment spell began in June, interaction with sex	-0.15	(0.06)	**
Unemployment spell began in July, interaction with sex	-0,14	(0,07)	***
Unemployment spell began in August, interaction with sex	-0,31	(0,08)	***
Unemployment spell began in September, interaction with sex	0,07	(0,07)	
Unemployment spell began in October, interaction with sex	-0,07	(0,08)	
Unemployment spell began in November, interaction with sex	-0,14	(0,06)	**
Number of months in unknown position in the 5 years before current, interaction with sex	0,00	(0,00)	**
No knowledge of Dutch, interaction with sex	0,19	(0,18)	
Number of months out of the labour market in the 10 years before current spell, interaction with sex	0,00	(0,00)	
Number of months out of the labour market in the 5 years before current spell, interaction with sex	0,00	(0,00)	
Educ. Attainment: professional bachelor, interaction with sex	0,03	(0,08)	
Drivers license car, interaction with sex	-0,06	(0,04)	

Table B.1, continued

Drivers license truck, interaction with sex	-0,11	(0,22)	
Drivers license bus, interaction with sex	-0,16	(0,20)	
Knowledge of Spanish, interaction with sex	0,04	(0,05)	
Country of birth: Turkey & Morocco, interaction with sex	0,12	(0,07)	*
Number of months in unemployment in the 10 years before current spell, interaction with sex	0,00	(0,00)	
Number of months in unemployment in the 2 years before current spell, interaction with sex	0,01	(0,00)	
Number of months of work in the 10 years before current spell, interaction with sex	0,00	(0,00)	
Number of months with sickness benefit in the 5 years before current spell, interaction with sex	0,00	(0,00)	
Having participated in a Dutch training course before current spell, interaction with sex	-0,22	(0,13)	*
Having participated in a training before current spell, interaction with sex	-0,10	(0,06)	*

Note: Standard errors are in brackets. *, **, *** indicate the precision of the estimate by showing whether the p-value of a two-sided significance test is below 10%, 5%, 1% respectively.

B.2 Tuning parameters of the MCF

As the MCF is a causal forest it has its usual tuning parameters. The important ones will be discussed in turn:

Variables for leaf splitting: 6, 15, or 40 randomly selected variables are used in each leaf splitting (out of about 61 that usually remained after feature selection described next). For most outcomes, 6 variables led to the smallest value of the objective function (which to be minimized) in the out-of-bag samples. Minimum leaf size for feature selection is set to 5.

Feature selection: A 20% random subsample was (exclusively) used to run a preliminary estimation of the MCF forest. The other 80% are used for the remaining estimation steps. The permutation based variable importance measure (VIM) was used to deselect variables. Since the standard one-variable-at-a-time VIM is always conditional on the other variables included, this was done in two steps: 1) VIMs are computed for all (91) variables. Then, these variables were sorted with respect to the VIMs and grouped into 10 groups. Next, groupwise VIMs (i.e. VIMs based permutating all variable in the group simultaneously) for these groups were computed. For all groups with a non-positive groupwise VIM (only), the following process is implemented to deleted variables. If the VIM of the worst group is non-positive, the variables in that group are deleted. If so, next a group-wise VIM for the worst and the second-to-worst group jointly is computed. If this is non-positive as well, the variables in the second-to-worst group are also deleted. This process is continued, until the group-wise VIM are positive. This process is computationally not cheap but it avoids deleting jointly variables that are relevant, but highly correlated (so that each is irrelevant given the others). This process leads to the removal of 30 variables for most outcomes.

The *minimum leaf size* was set to 5.

The number of *subsampling replications* was set to 1000 and *67% of the observation used to build the forest are contained in each subsample* (this comparatively high number was chosen because although the overall number of observations is large, the number of observations in each programme and nonparticipation (that eventually determines the depth of the splitting) is rather low.

B.3 Tuning parameters of the Post-Lasso for the IATEs

The penalty term of the LASSO was determined by a grid of 100 different values, starting with the model without covariates to the model with all covariates. On this grid the optimal value was determined by ten-fold cross-validation using the mean squared prediction error of the Post-Lasso.

Appendix C: Additional results

C.1 Average treatment effect on the treated

Table C.1 shows how the effects differ across the populations of programme participants for all average population effects. Numbers in bold always relate to the effect of the programme for its own population of participants (ATET). If caseworkers maximise effects, then one should expect the bold number to be the largest entry in each row.

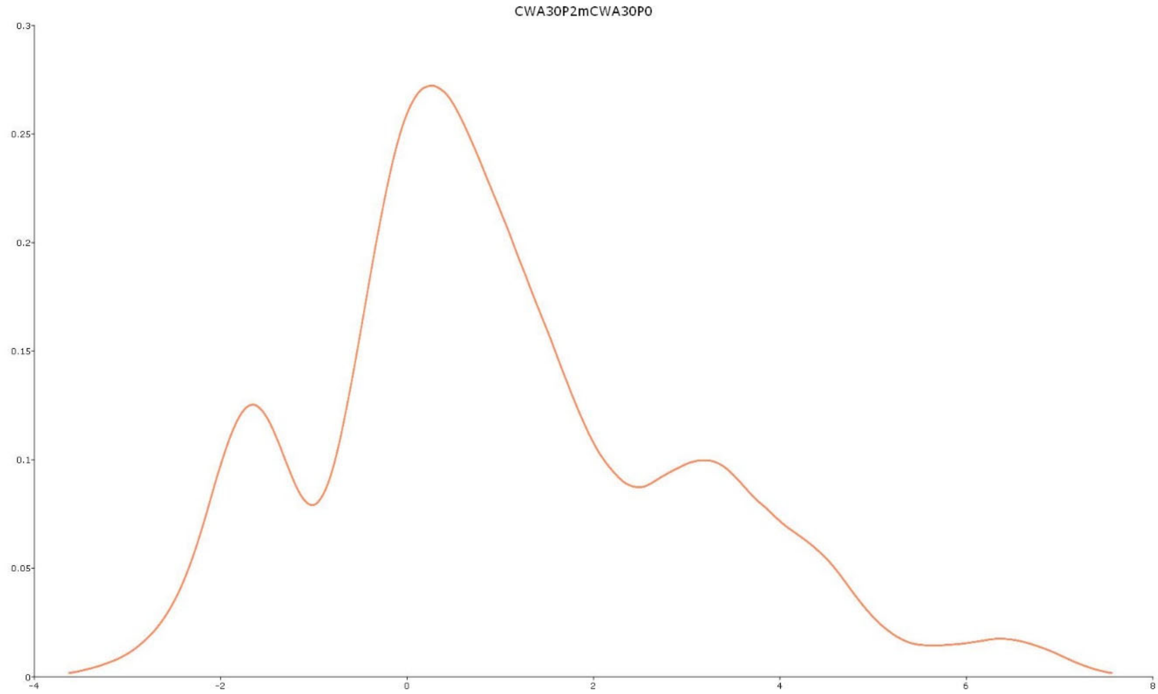
Table C.1: Comparison of the effects for the different programmes on cumulative months in employment, unemployment and out of the labour force for different populations by participation status

	No ALMP participation (NOP)	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
Cumulative months in employment 9 months after programme start				
SVT – NOP	0.1 (0.2)	**0.1 (0.1)**	-0.0 (0.2)	0.0 (0.2)
LVT – NOP	-1.1 (0.1) ***	-1.1 (0.1) ***	**-1.1 (0.1)** ***	-1.1 (0.1) ***
OT – NOP	-1.6 (0.1) ***	-1.5 (0.1) ***	-1.6 (0.1) ***	**-1.6 (0.1)** ***
LVT – SVT	-1.2 (0.2) ***	-1.2 (0.2) ***	**-1.1 (0.2)** ***	-1.2 (0.2) ***
OT – SVT	-1.6 (0.2) ***	-1.6 (0.2) ***	-1.6 (0.2) ***	**-1.6 (0.2)** ***
OT – LVT	-0.4 (0.2) **	-0.4 (0.2) **	-0.5 (0.2) **	**-0.4 (0.2)** **
Cumulative months in *employment between month 22 and month 30* month after programme start				
SVT – NOP	1.4 (0.2) ***	**1.4** (0.2) ***	1.3 (0.2) ***	1.3 (0.2) ***
LVT – NOP	1.2 (0.2) ***	1.1 (0.2) ***	**1.0** (0.2) ***	1.1 (0.2) ***
OT – NOP	0.5 (0.2) **	0.5 (0.2) **	0.4 (0.2) **	**0.4 (0.2)** **
LVT – SVT	-0.1 (0.3)	-0.3 (0.2)	**-0.2** (0.3)	-0.2 (0.3)
OT – SVT	-0.9 (0.3) ***	-1.0 (0.3) ***	-0.9 (0.3) ***	**-0.9 (0.3)** ***
OT – LVT	-0.8 (0.3)	-0.7 (0.3)	-0.6 (0.2)	**-0.7 (0.2)**
Cumulative months in employment 30 months after programme start				
SVT – NOP	3.4 (0.5) ***	**3.4 (0.4)** ***	2.8 (0.6) ***	3.1 (0.6) ***
LVT – NOP	1.0 (0.5) **	0.8 (0.5)	**0.4 (0.4)**	0.6 (0.4)
OT – NOP	-1.4 (0.5) ***	-1.5 (0.5) **	-1.8 (0.5) ***	**-1.7 (0.5)** ***
LVT – SVT	-2.4 (0.7) ***	-2.6 (0.6) ***	**-2.4 (0.7)** ***	-2.4 (0.7) ***
OT – SVT	-4.8 (0.7) ***	-4.9 (0.6) ***	-4.7 (0.7) ***	**-4.8 (0.7)** ***
OT – LVT	-2.4 (0.7) ***	-2.3 (0.7) ***	-2.3 (0.6) ****	**-2.3 (0.6)** ***
Cumulative months in unemployment 30 months after programme start				
SVT – NOP	-1.9 (0.3) ***	**-2.0** (0.3) ***	-1.7 (0.3) ***	1.8 (0.3) ***
LVT – NOP	0.9 (0.4) **	0.8 (0.5) *	**1.1** (0.4) ***	1.0 (0.4) ***
OT – NOP	2.7 (0.5) ***	2.6 (0.5) ***	2.9 (0.4) ***	**2.8 (0.4)** ***
LVT – SVT	2.8 (0.5) ***	2.8 (0.6) ***	**2.9** (0.5) ***	2.8 (0.5) ***
OT – SVT	4.6 (0.6) ***	4.6 (0.6) ***	4.6 (0.5) ***	**4.6 (0.5)** ***
OT – LVT	1.8 (0.6) ***	1.8 (0.7) ***	1.8 (0.6) ***	**1.8** (0.6) ***
Cumulative months out-of-the-labour force 30 months after programme start				
SVT – NOP	-1.5 (0.3) ***	**-1.4 (0.3)** ***	-1.3 (0.2) ***	-1.4 (0.3) ***
LVT – NOP	-1.9 (0.3) ***	-1.7 (0.2) ***	**-1.6 (0.2)** ***	-1.7 (0.2) ***
OT – NOP	-1.4 (0.3) ***	-1.3 (0.3) ***	-1.1 (0.3) ***	**-1.2 (0.3)** ***
LVT – SVT	-0.4 (0.4)	-0.2 (0.3)	**-0.4 (0.3)**	-0.4 (0.3)
OT – SVT	0.1 (0.3)	0.2 (0.3)	0.2 (0.3)	**0.1 (0.3)**
OT – LVT	0.5 (0.4)	0.4 (0.3)	0.5 (0.3) *	**0.5 (0.3)**

Note: Outcomes measured in months. ATET for the particular programme in bold. All effects are population averages for the respective programme participants given in the column. Standard errors are in brackets. *, **, *** indicate the precision of the estimate by indicate whether the p-value of a two-sided significance test is below 10%, 5%, 1% respectively.

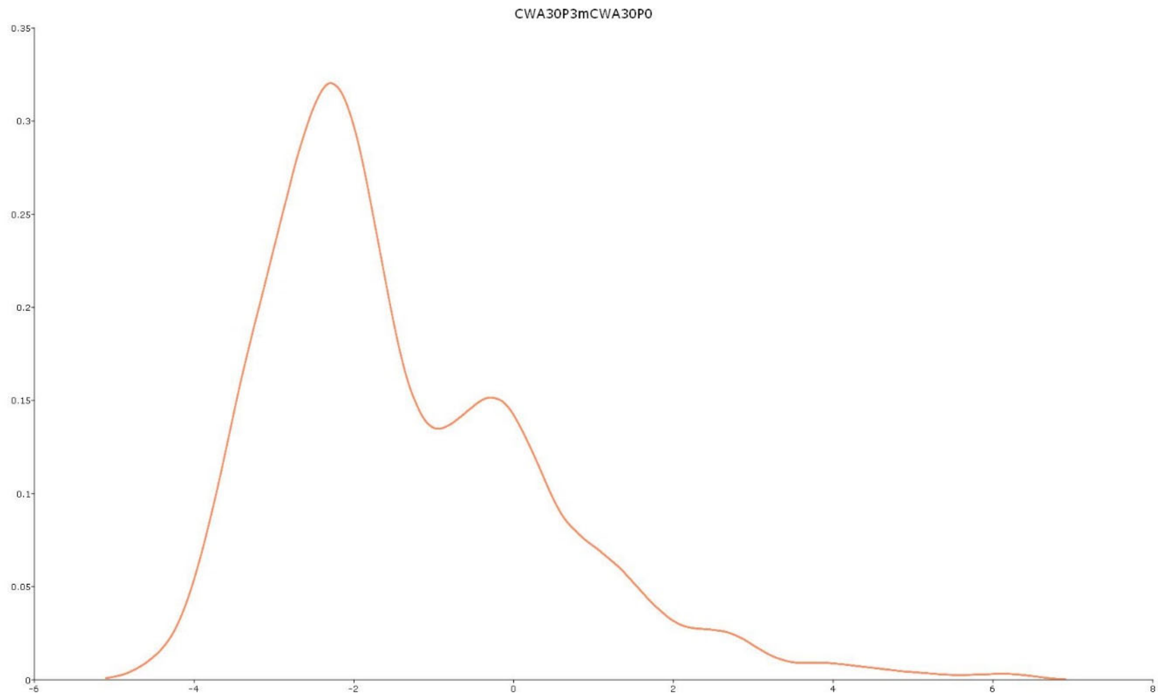
C.2 Distribution of estimated IATEs

Figure C.1: Distribution of estimated IATE of LVT vs. NOP



Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.

Figure C.2: Distribution of estimated IATE of OT vs. NOP



Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.