CESifo
AREA
CONFERENCES 2020

Economics of Education

Online, 30 November–1 December 2020

Priority to Unemployed Immigrants? A Causal Machine Learning Evaluation of Training in Belgium

Bart Cockx, Michael Lechner, and Joost Bollens

CESifo

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

Bart Cockx[$], Michael Lechner[$$], Joost Bollens[$$$*]

[$] Department of Economics, Ghent University
[$$] Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen
[$$$] Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding (VDAB)

This version: September 2020

Date this version has been printed: ***29 September 2020***

Abstract: Based on administrative data of unemployed in Belgium, we estimate the labour market effects of three training programmes at various aggregation levels using Modified Causal Forests, a causal machine learning estimator. While all programmes have positive effects after the lock-in period, we find substantial heterogeneity across programmes and unemployed. Simulations show that "black-box" rules that reassign unemployed to programmes that maximise estimated individual gains can considerably improve effectiveness: up to 20% more (less) time spent in (un)employment within a 30 months window. A shallow policy tree delivers a simple rule that realizes about 70% of this gain.

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

JEL classification: C21, J68.

Addresses for correspondence:
Bart Cockx, Ghent University, Department of Economics, Sint-Pietersplein 6, 9000 Gent, Belgium, bart.cockx@ugent.be, sites.google.com/site/bartcockxsite.
Michael Lechner, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstr. 14, 9000 St. Gallen Switzerland, Michael.Lechner@unisg.ch, www.michael-lechner.eu, www.researchgate.net/profile/Michael_Lechner.
Joost Bollens, VDAB, Keizerslaan 11, 1000 Brussels, Belgium, joost.bollens@vdab.be.

[*] Bart Cockx is also affiliated with IZA Bonn, CESIfo, Munich, IRES, Université catholique de Louvain and ROA, Maastricht University. 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). We thank Michael Knaus, Anthony Strittmatter and Gabriel Okasa for carefully reading and commenting on a previous draft of the paper. Previous versions of the paper were presented at research seminars of the LMU University of Munich, the University of Milan, Competence Centre of Microeconomic Evaluation of the Joint Research Centre of the European Commission, and at the IMT in Lucca, as well as in workshops and conferences in Copenhagen, St Gallen, Engelberg, and Basel. We thank participants, in particular Bas Van der Klaauw, Claudio Lucifora, Regina Riphahn, Claus Schnabel, and Andreas Steinmayr, for helpful comments and suggestions. We thank the VDAB for providing us access to the data. The VDAB stocks the data and has engaged to make the data available to researchers who wish to replicate the current research. The usual disclaimer applies.

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 whom. 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 to which 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 uncover the underlying heterogeneity of the causal effects, a goal for which traditional econometric methods fail to provide a systematic solution. At the same time, we will show that the machine learning methods will provide very similar and, hence, reliable estimates of the *average* treatment effects as the traditional methods.

As mentioned , in this paper identification of the causal effect relies on the assumption of unconfoundedness. Knaus, Lechner and Strittmatter (2018) evaluated 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 were taken from the Swiss social security records that were used to evaluate a job search programme for the unemployed (Knaus, Lechner and Strittmatter 2017). Knaus, Lechner and Strittmatter (2018) conclude that the Causal Forest based ML methods, particularly the Generalized Forest by Athey, Tibshirani, and Wager (2018), belong to the best performing estimators if explicitly adjusted – in another step – 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 within a single step by penalizing tree splits that do not reduce selection bias. 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 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 administrative data from 2003. They find substantial effect heterogeneity, but only during the first 6 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.

Unlike aforementioned papers, this paper considers multiple programmes and explicitly exploits the estimated effect heterogeneity to propose reallocation rules that could increase the programme performance. Two approaches are considered. A first approach aims at finding a rule that reassigns the unemployed to programmes based on the estimated individual programme effects such that it maximizes for the population of entrants into unemployment a chosen objective – here, with equal weight, respectively maximizing and minimizing the time spent in employment and unemployment over a 30 months horizon. The disadvantage of this "black-box" approach beyond some concerns about 'overfitting' is that it is not transparent, so that caseworkers may be reluctant to implement such rules (e.g. Lechner and Smith, 2007) and/or it may conceal that the rule implicitly raises ethical or societal concerns (e.g. Whittlestone, Nyrup, Alexandrova, Dihal, and Cave, 2019; Reddy, Allan, Coghlan, and Cooper, 2020). We therefore also consider an approach recently proposed by Zhou, Athey, and Wager (2019) in which optimal policy rules are derived using shallow decision trees with a small number of nodes, resulting in simple rules that are easy to understand and judge.

The analysis uses administrative data of the Flemish PES and 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 9 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 1991. This makes the assumption of unconfoundedness, which is used for identification, arguably plausible (see, e.g., Lechner, Miquel and Wunsch 2011, and Biewen, Fitzenberger, Osikominu and Paul, 2014). 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 well as in the medium run (post-programme effects): SVT clearly

[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.

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. 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, education, 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 simulations that change the allocation of programme participants according to the individualized effectiveness of these programmes, following the two aforementioned approaches. Based on the black-box approach, we find that a reallocation that keeps programme capacity fixed can increase the time in employment within a 30-months-time-window by nearly 20% and reduce the time in unemployment by 8%. A policy tree of depth 4 defines a policy rule that can achieve about 80% of this gain, while a policy tree of depth 3 results in a considerably simpler rule with similar employment gains, but at the cost of reducing the time spent in unemployment only by about 4%, i.e. half of the potential gain. The policy rule of depth 3 reveals that most employment gains are achieved by reallocating workers

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.

older than 28 years, born in an Eastern EU country to vocational training[3] and those originating from another country with poor proficiency in Dutch (the official language in Flanders) to OT. This confirms that with regards to employment the returns to programme participation are highest for recent immigrants.

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, including the placebo analysis, and ends by concluding remarks. An appendix provides additional details on the data, the estimator used, the results, and the sensitivity analysis.

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 (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

[3] Those with more than 9 months of employment experience in Belgium to LVT and the others to SVT.

transmission by the regional PES, these tasks have been transferred to the regional PES since 2016 (in Brussels since 2017).

In Belgium unemployed workers are 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 after a waiting period of one year. Second, workers with recent experience of at least one year in salaried employment qualify for UB after involuntary lay-off.[4] We focus in this paper on the latter scheme. In view of our finding that programme participation is most effective for recent immigrants, it is important to note that until October 2016 (relevant period for the empirical analysis) one could qualify for unemployment benefits in Belgium based on proof of equally long periods of salaried employment abroad as long as at least one day was worked in Belgium.[5]

Unlike in other countries, the benefits are paid out without time limit. The UB are 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 UB 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.[6]

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.

[4] Workers older than 35 must have worked more than one year to qualify for UB.

[5] See RVA/ONEM in Dutch, French or German (accessed on 2020-04-27).

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

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 must 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 rights and duties. During 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* (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 include 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 add up 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 for follow-up of participants over 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 need not 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 9 months of unemployment, i.e.

the not yet treated group (Sianesi, 2004); (2) 3,640 individuals who started within the first 9 months an ALMP retained in the analysis; (3) 13,618 individuals who entered within the first 9 months an ALMP and who are not retained for the evaluation, because (i) the placebo tests suggested concerns about selection bias for the evaluation of intensive counselling, (ii) almost only foreigners with limited language skills participated in Dutch language training, 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.

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: Retained individuals are aged between 21 and 55 years at registration and started claiming UB 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 lasted 8.82 months on average.

Programme participants are classified according to the first programme they participate in. The programmes considered in the analysis are the following. First, following Lechner, Miquel and Wunsch (2011), 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. We stress that these programmes do not only differ by duration, but also by content. In fact SVT and LVT both aggregate training for different professions in different sectors. LVT is therefore not just more of SVT.

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.

3.3 Confounding and outcome variables

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 usually categorical variables would be transformed into dummy variables in a regression-type setting,[7] this would correspond with 175 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

[7] 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).

years, the ALMP participation history during previous unemployment spells, information about the job seeker's job preferences and the corresponding professional experience, the calendar 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)[8] in the unemployment spell (maximum 274 days).

For a correct interpretation of the our findings reported below, the reader should be aware of the following data limitations: (i) labour market history is recorded only as from the first entry in unemployment, which means that those experiencing a first unemployment spell will have no reported work experience; (ii) as mentioned in Section 2, foreign work experience counts to qualify for UB, but as it is not measured in the data, the data contain individuals – in particular recent immigrants – with less than one year of employment experience even if this is less than what is required to qualify for UB. This is a sizeable fraction of the selected sample: about 30% report less than 12 months of employment experience in the last two years and about 13% no experience at all. However, by combining the information on age and educational attainment, the data implicitly can proxy the employment history of natives who experience a first unemployment spell, because this group has most likely been uninterruptedly employed since leaving education. This data limitation is therefore essentially an issue of interpretation. The missing information about the employment history of recent immigrants is not necessarily problematic, because (i) foreign employment is less valuable than local labour market experience and (ii) the combination of education, age and other socio-demographic information might proxy the missing information quite well. In the placebo analysis reported in Section 7.1 we find no evidence of specific violations for the non-native population of the conditional independence assumption on which the estimator relies (Section 4.2).

[8] More on this in Section 4.4.

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 %) 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.1. 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.

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 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.

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	56324	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).

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.

In Appendix A.2 we further analyse the direction of the selection bias between the different programmes. From this analysis it can be clearly concluded that participants in SVT are less qualified for the labour market than those in LVT, while participants in LVT and OT have a similar profile. We will refer to this evidence when discussing the relative average effectiveness of these programmes in Section 5.1.

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, $y_i = \sum_{d=0}^{3} \underline{1}(d_i = d) y_i^d$ ($\underline{1}(\cdot)$ denotes the indicator function, which is one if its argument is true and zero otherwise).[9] 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:

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

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

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

9 If not obvious otherwise, capital letters denote random variables, and small letters 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):

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

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 (with the qualifications mentioned in Section 3.2), 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 to 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 split 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.[10] Lechner (2018) develops these ideas further by improving on the splitting rule for the individual trees, penalizing splits that do not reduce selection bias, 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-

[10] Athey, Tibshirani, and Wager (2019) 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 a 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 that 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 downwards, 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 insights, 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.[11] 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

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

suggested by Lechner (1999, 2002), to accommodate for the critique of Fredriksson and Johansson (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.

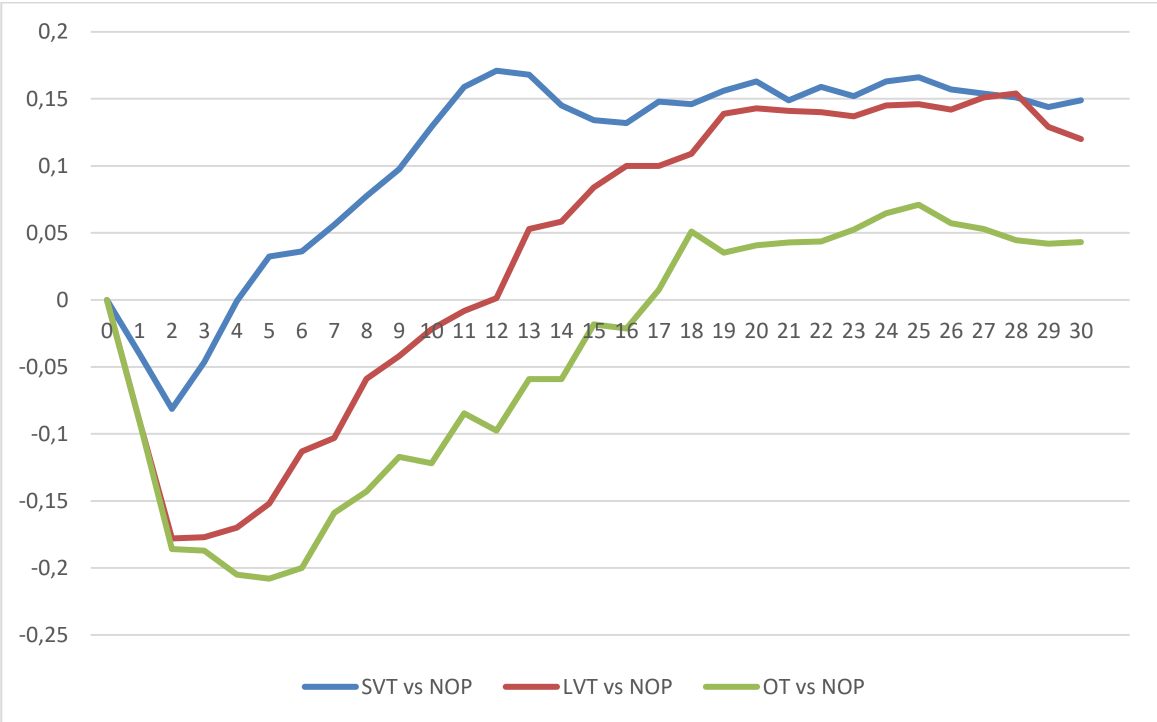
Then, for the arguably most important short- and medium-run outcome, namely employment, we investigate more thoroughly the heterogeneities with respect to the programmes and groups of unemployed by their programme participation. Subsequently, the heterogeneity of the most policy relevant medium-run effects 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 employment probability 2 to 30 months after programme start



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.[12] 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-

[12] 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), particularly 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. However, Lechner, Miquel and Wunsch (2011), who also find a similar pattern for the employment effects when comparing long to short training programmes in Germany, do not find that longer training increase earnings more. Consequently, as in the German study, the finding that the longer training is not more effective is more likely related to the differences in the training content (see Section 3.2). This interpretation is also more plausible than that the effect of SVT is biased upwards due to a failure of the unconfoundedness assumption. On the one hand, such a failure is not consistent with the findings of the placebo analysis that we will report in Section 7.1. On the other hand, Altonji, Elder and Taber (2005) provide plausible theoretical arguments that, in case of a bias on unobservables, the direction of this bias should be the same as the one on observables. This would then rather, if any, suggest a downward bias of the effects of SVT, as in Section 3.3 we presented evidence that participants in SVT are disadvantaged in terms of employability relative to those in LVT.

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. Even if one is not convinced by the absence of placebo effects reported in Section 7.1, using again the aforementioned argument of Altonji, Elder and Taber (2005), the finding in Section 3.3 that participants in OT have a comparable profile as those in LVT makes it unlikely that the smaller effects of OT would be caused by a violation of the unconfoundedness assumption. 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 measures (months in unemployment, months out-of-the-labour-force).

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 increases 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.

Table 5.1: Effects for the different 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 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) **	**2.0 (0.1)**
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.

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 at conventional significance levels. However, the reason for this rejection is that the ATET is worse than the ATE.

This means that either the treatment effects are fairly homogeneous for these programmes or that caseworkers' assignment to the different programmes is close to being 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. In Section 6 we discuss the gains that the PES could make by improving the assignments of unemployed to the different programmes.

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 $\chi^2(3)$. *, **, *** indicate whether the p-value is below 10%, 5%, 1% respectively.

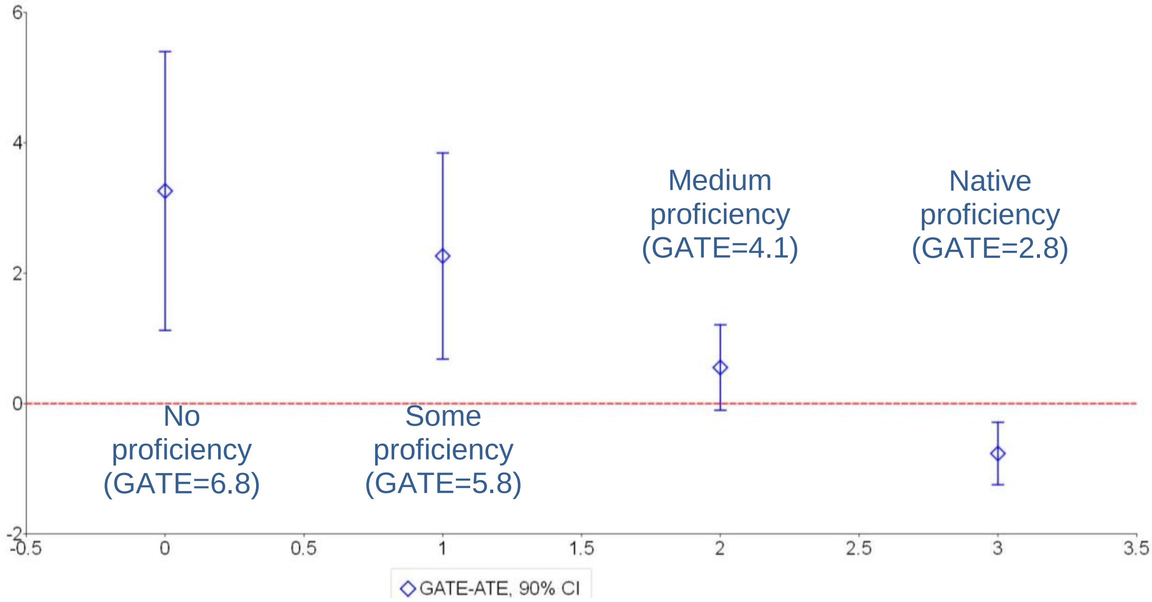
5.2 Heterogeneity with respect to policy relevant variables

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, 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 they lost their job. Recent unemployment history helps therefore identifying a population that is more loosely attached to the labour market. 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 (Piton and Rycx, 2020). 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.

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 two dimensions (proficiency in Dutch and country of birth), 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%).

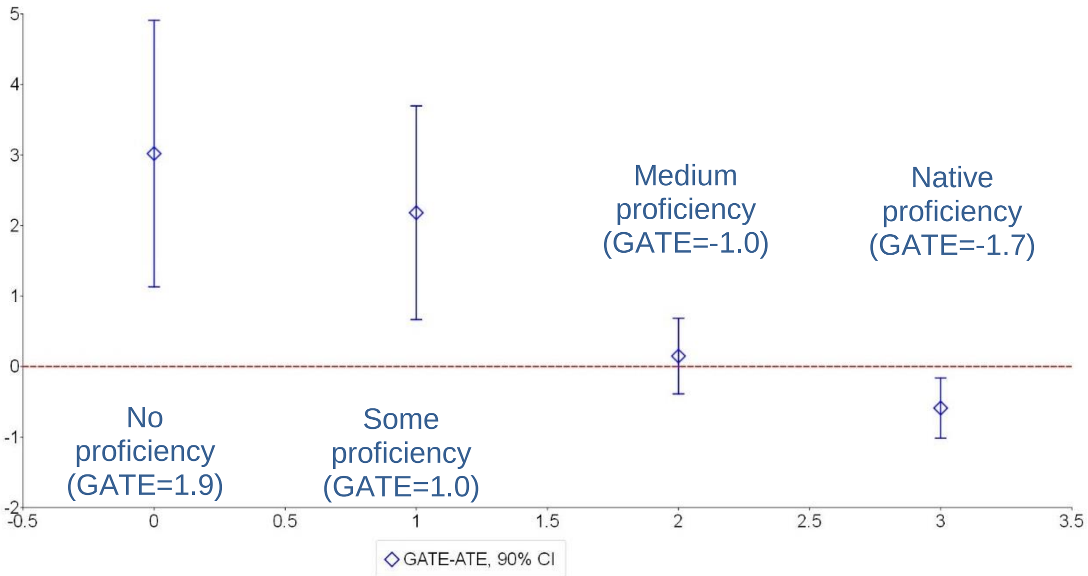
Figure 5.2: Difference of GATEs to ATE of SVT relative to NOP for the four proficiency levels in Dutch – Cumulative number of months employed 30 months after programme start



Note: Dutch proficiency displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. Dutch proficiency varies between no proficiency (0) and native proficiency (3).

The finding that the effectiveness of the programme decreases with language proficiency may seem contradictory, as a minimal requirement for training to be effective is that one can understand the instructor. It appears, however, that the PES automatically provides language assistance to participants who have insufficient understanding of Dutch.

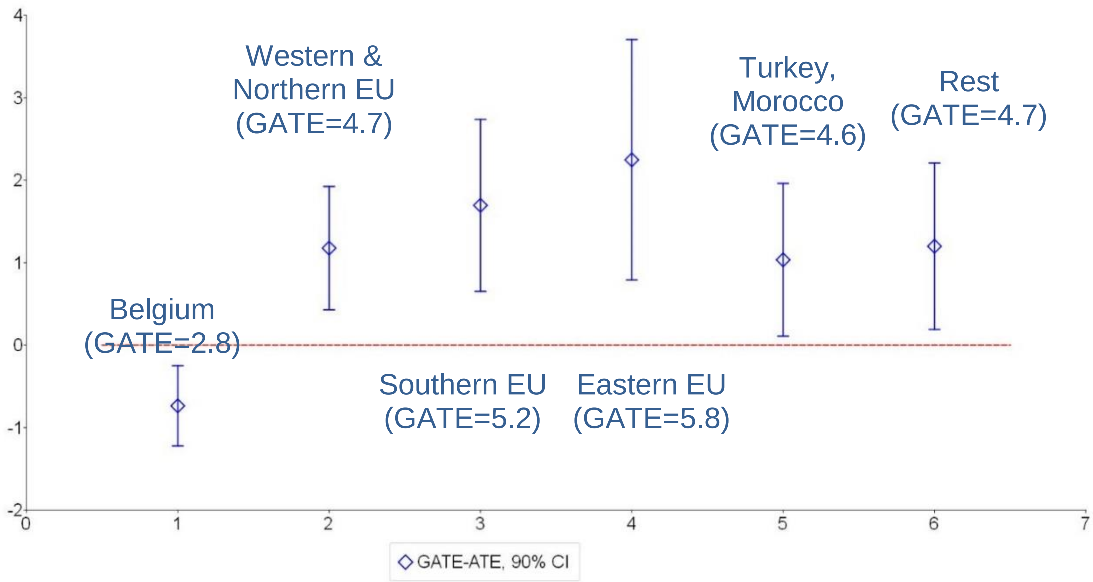
Figure 5.3: Difference of GATEs to ATE of OT relative to NOP for the four proficiency levels in Dutch – Cumulative number of months employed 30 months after programme start



Note: Dutch proficiency displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. Dutch proficiency varies between no proficiency (0) and native proficiency (3).

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 5.4: 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 displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. 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 Eastern European Union countries (5.8 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, remain significantly higher (4.6 months) than for Belgians (2.8 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 (significantly different at 5%). Together with the previous finding on proficiency in Dutch this strongly suggests that SVT is more effective for migrants who recently migrated to Belgium.

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 measured 30 months after the programme start. 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 low educated, those living in a city, and 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.[13]

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

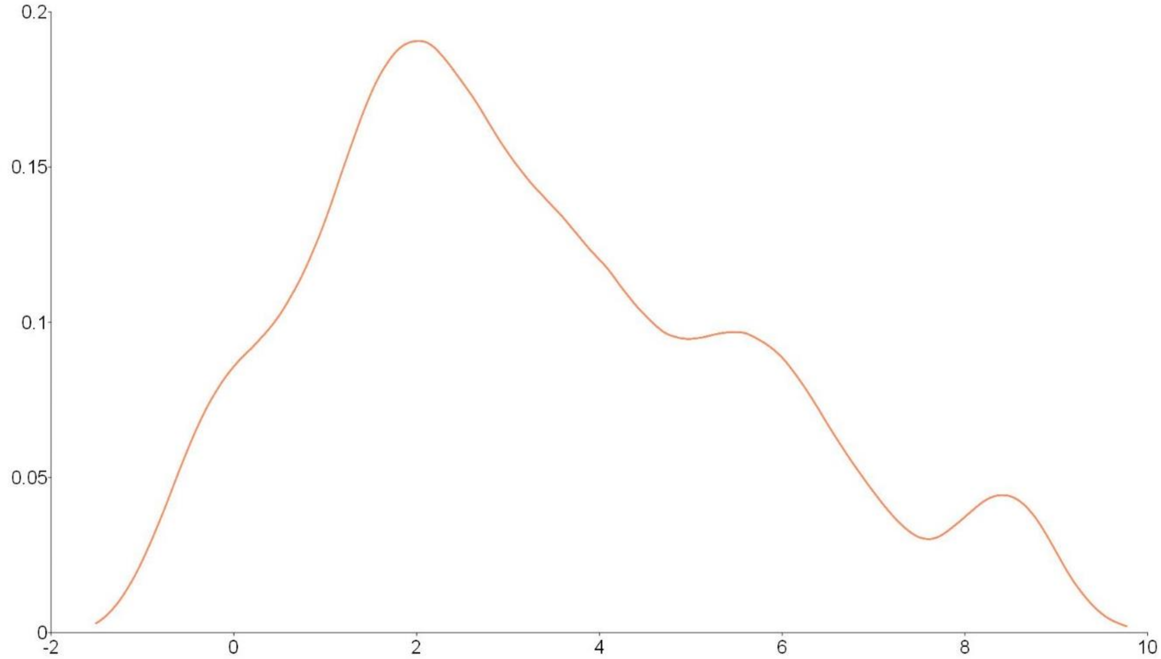
In this section, we present the results for the individualized average effects (IATEs), which present the finest level of granularity available. To avoid flooding the reader with numbers, we will concentrate on the cumulative medium-term employment outcomes for the comparison to NOP (no programme participation) which are likely to be the most policy relevant. We first describe the extent of heterogeneity in the programme effects. We present the results of a k-means clustering analysis to get an informal characterization of sub-groups clustered 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. The mean of these effects is 3.4 month (as shown in Table 5.1) and the standard deviation 2.2. About 34% of the estimated IATEs is 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

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

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



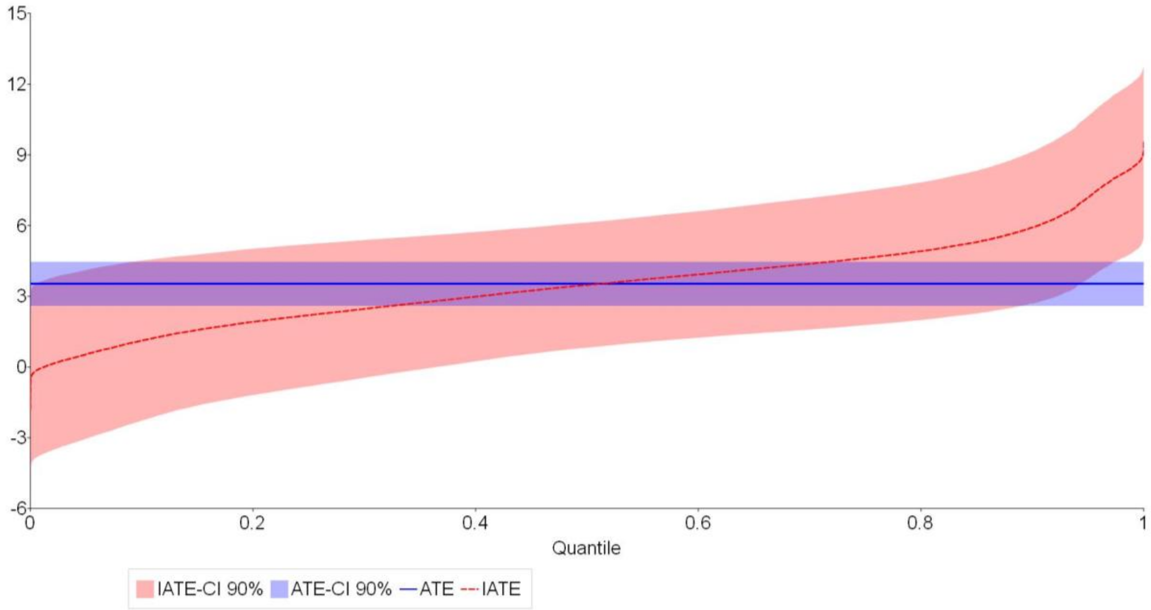
Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.

The respective figures for the other programmes, as well as the sorted effects (and inference) of the difference of the IATE to the ATE show qualitatively similar patters like the ones presented here and are thus moved to Appendix C.2.

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. 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, 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 8 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.

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 of IATEs based on estimated standard errors and normal distribution. Standard errors are smoothed by Nadaraya-Watson regression (Epanechnikov kernel with Silverman bandwidth).

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 5 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 Turkey and Morocco and from the rest of the world

– migrants with very bad labour market performance in Flanders[14] – are most represented in the second most effective group. Taken this together with that fact that the most effective clusters comprise individuals with the least recent and less recent employment and unemployment experience (and that the mean age rules out recent school leavers), it becomes obvious that the group for which the programmes are most effective consist mainly of recent migrants. Additionally, living in a city rather than in rural areas and postponed programme starts are also associated with larger programme effects.

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 8 clusters. Covariates are not used to form clusters. K-means ++ algorithm used (Vassilvitskii, 2007).

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

[14] 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). See also Piton and Rycx (2020) for more evidence.

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.

Finally, one way to characterise the employability of the unemployed is to consider their estimated months of employment without the programme (NOP). The last row in Table 5.3 shows that in those 8 groups, programme effectiveness monotonically decreases with employability, which is consistent with the picture of heterogeneity revealed so far.

6 Policy simulations

So far, we have documented considerable heterogeneity in the effectiveness of the programmes. Do caseworkers exploit this information in the sense of assigning the unemployed to the programmes that work best for them, and, if not, to what extent could a different assignment improve the performance of the PES? We shed some light on the answers to these questions by simulating hypothetical programme allocations which we compare to the observed allocation. For generating the new assignments, we use two approaches. The first one is a "black-box" approach based on individually allocating the treatment with the highest estimated potential outcome. However, such an approach may have the problem that caseworkers will not trust and thus not follow such rules (when made available by an AI system), as they cannot be expected to understand how they came about. Recently, Zhou, Athey, and Wager (2019) fill this gap by proposing a method – valid in the context of multiple treatments – that allows to derive policy rules based on decision trees. When restricting this method to shallow trees with a small number of nodes, the resulting rule will be simple and easy to understand and implement, or, alternatively, reveal ethical or societal concerns with the rule (e.g. Whittlestone, Nyrup, Alexandrova, Dihal, and Cave, 2019; Reddy, Allan, Coghlan, and Cooper, 2020). Thus, caseworkers will be more likely to follow it, at least if they do not strongly disagree. However, a potential drawback is that a shallow tree might not approximate the optimal policy as well as a black-box AI recommender system.

The two first lines of Table 6.1 contain the observed allocation and a randomised allocation. The next 5 lines presents the black-box rules, while the lower panels of the table show the allocations resulting from decision trees of depth three and four. The first column of Table 6.1 features the description of the allocation, followed by the shares of the population allocated to the different programmes (in %), and the population average effects of the policies on the outcomes of interest, namely the average number of months employed (Emp), the average number of months unemployed (UE), and the average number of months out-of-the-labour force (OLF). These three outcomes add up to 30. We assume throughout that the policy rules aim at maximizing, respectively minimizing, with equal weight, the number of months in employment, respectively unemployment. The last two columns show the relative change (in %) in employment and unemployment for the subpopulation of those individuals who ended up in another treatment state than the one observed.

6.1 Comparing observed to random assignment

Comparing the observed allocation that serves as a reference point for the simulations with complete random assignment (with probabilities of assignment equal to the observed shares of participation in the different programmes) shows that the average programme impacts of the two allocations are very similar. The comparison of the outcome of the random with the observed allocation suggests that caseworkers are unlikely to base their assignment decisions on the effect heterogeneities described in this paper. This appears to be in line with official policy. The PES in Flanders used job placement rates, like “70% of the participants must be employed within 6 months after the end of the training", as targets for evaluating caseworkers instead of effect targets. Thus, the Flemish PES can improve the performance of the training programmes by adjusting the assignment according to the expected programme performance of individuals as estimated by their IATE.

6.2 Black-box assignment rules

Next, we consider several black-box assignment schemes that depend on the available programme capacity and the degree of certainty about the effectiveness of programmes and some priority rule. The third line (*no constraint*) shows the results in case without any constraints. 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 from 16.1 to 19.4 (an increase of more than 20%), reduce the number of months in UE from 10.9 to 8.9 (a reduction by about 18%) and in OLF from 3.0 to 1.6 (a reduction by 53%). Obviously, this would be very costly, because it would massively increase the number of SVT participants. Whether the overall benefits would outweigh these costs is unclear, because we do not have access to the information, such as detailed programme costs, necessary to conduct such cost-benefit analysis.

Table 6.1: Overall effects of some simulated hypothetical programme allocations

	Share of different programmes in %			Cumulative # of months 30 months after programme start			Gain for switchers in %	
	SVT	LVT	OT	Emp	UE	OLF	Emp	UE
Observed	2.1	2.0	1.8	16.1	10.9	3.0	-	-
Random	2.0	1.9	1.9	16.1	10.9	3.0	0.9	-0.1
Black-box – no constraint	97.3	0.2	0.2	19.4	8.9	1.6	21.9	-18.5
Black-box – no constraint, only significant	58.1	1.6	0.5	18.8	9.4	1.8	33.4	-23.0
Black-box – constrained, preference to largest gains	2.1	2.0	1.8	16.4	10.8	2.8	13.6	-6.3
Black-box – constrained, sequential optimization	2.1	2.0	1.8	16.4	10.8	2.9	19.3	-8.0
Black-box – constrained, preference to lots of past UE*)	2.1	2.0	1.8	16.2	10.9	2.9	6.6	-2.5
Simple - Policy tree 3 level, constrained	2.0	2.0	1.8	16.3	10.8	2.9	15.0	-4.2
Simple - Policy tree 4 level, constrained	2.1	2.0	1.8	16.4	10.8	2.8	14.6	-6.7

Note: Emp: Employed; OFL: Out-of-labour-force; UE: Unemployed. Allocations minimize unemployment and maximise employment (equally weighted). *) If programmes capacity becomes a constraint, preference is given to the highest number of months in unemployment over the last 10 years. *Sequential optimisation* means that starting with the observed allocation, programme states of any pair of individuals are pairwise switched if overall the outcome is improved and the budget constraint is kept.

The previous allocation ignores the fact that some estimated IATE are positive (or reverse) just because of estimation error. Therefore, in line four (*no constraint, only significant*) we report the outcome of a simulation in which we assign individuals only to programmes if the corresponding IATEs are significantly positive or negative at the 2.5% level of a one-sided statistical test in terms of their effect on the time spent in, respectively, 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 population average programme effects decrease in a much lesser proportion. Nevertheless, almost 60% of the unemployed are allocated to SVT. The average gain in terms of employment is still in total 2.7 months and corresponds to a 33% improvement for those observations reallocated.

The next scenarios consider cases in which it is assumed that the training capacity of the PES is constrained to the observed one (and thus programme costs remain approximately the same[15]). Since only 6% of the population participates in training, this cannot dramatically reduce the overall gains that a relocation can generate. However, for those actually affected by the reallocations, the gains may be very substantial. Three scenarios are considered. In the first scenario, priority is given to individuals with the highest returns to programme participation (*constrained, preference to largest gains*).[16] The average employment gain for the population is still 9.3 days (i.e. 0.31 months). Per reallocated unemployed this corresponds to a substantial months-in-employment gain of about 14%. The next scenario tries to get close to a constrained optimum by finding pairs of observations for which exchanging the treatment status improves the employment, resp. UE with equal weight. Such exchanges are computed sequentially until

[15] Unfortunately, average costs of the programme are not available in the data. Therefore, we use programme shares to approximate a budget constraint.

[16] 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 largest. 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.

no such pairs can be found. This leads to a substantial gain in terms of months of employment of about 19% and a reduction in months of unemployment of about 8% for those who were reallocated, an improvement of respectively 42% and 21% relative to the previous scenario. In the last black-box approach we use a priority rule that is based on other criteria than programme effectiveness to deal with capacity constraints. This can be justified by concerns for equity or affirmative action, or by political constraints. Concretely, we first rank programmes for each individual according to the "best" estimated counterfactual outcome. If this leads to fewer participants than programme slots, then all these individuals are assigned to this programme. If not, individuals are assigned according to some priority rule, and not assigned individuals to their next best programme. This continues until all programme slots are filled. In Table 6.1 priority is given to individuals with the most UE experience in the last 10 years. Relative to the previous rule, this reduces the relative gain of the switchers to about one third. This is in line with expectations, because we have found that the programme effectiveness is highest for those with the *least* UE experience (Table 5.3). In Appendix D.4 some other priority rules are considered: preference to the unemployed with the worst NOP, to those with poor proficiency in Dutch and to "recent migrants".

6.3 Assignment rules based on shallow decision trees

We now take up the proposal by Zhou, Athey, and Wager (2019) and use shallow decision trees of depth 3 (resulting in 8 strata with potentially different treatment allocations) and 4 (16 strata) to obtain a more intuitive, interpretable rules. The unconstrained allocations are obtained by a slight modification of Algorithm 2 of Zhou, Athey and Wager (2019) which should lead to some better allocations at the expense of somewhat higher computational costs. In the original paper it is not discussed how to deal with categorical variables or with constraints which are both important issues in this setting. The details of these aspects are documented and explained in Appendix B.4. The case without constraints is not discussed, because it assigns virtually all unemployed to SVT and is therefore not interesting.

Table 6.2: Assignment rules of shallow decision trees

Tree Depth = 3				
Age ≤ 28.2	Age > 28.2			
	Country of birth (1, 2, 3, 5, 6)		Country of birth (4)	
	No proficiency in Dutch (0)	At least some proficiency in Dutch (1, 2 3)	Worked in Belgium ≤ 9 months in last 2 years(*)	Worked in Belgium > 9 months in last 2 years
NP	OT	NP	SVT	LVT

Tree Depth = 4										
Age ≤ 34.9				Age > 34.9						
Poor Dutch proficiency (0, 1, 2)			Profi-cient in Dutch (3)	Education (2, 5, 6, 10, 11, 13)			Education (1, 3, 4, 7, 8, 9, 12)			
No empl. in Belgium in last 2 years (*)	One month or more empl. in Belgium in last 2 years (*)			Country of birth (1, 2, 3, 5)	Country (4, 6)		Country (1, 3, 5)		Country (2, 4, 6)	
	Country of birth (1, 3, 5, 6)	Coun-tryof birth (2, 4)			Start UI after 12/2014	Start UI in 12/2014	Poor Dutch profi-ciency (0,1)	Profi-cient in Dutch (2, 3)	No empl. in Bel-gium in last 2 years(*)	1 month or more empl. in Belgium in last 2 years (*)
NP	NP	LVT	NP	NP	NP	LVT	OT	NP	SVT	NP

Note: In the table strata leading to the same allocation are merged as to simplify the presentation. Country of birth: 1 = Belgium, 2 = Western & Northern EU, 3 = Southern EU, 4 = Eastern EU, 5 = Turkey & Morocco, 6 = rest of the world; Proficiency in Dutch: 0 = no, 1 = some, 2 = medium, 3 = native; Education: 1 = high school (HS) drop-out (after 1st level), 2 = part-time HS degree, vocational track, 3 = HS drop-out (after 2nd level) general or artistic track, 4 = HS drop-out (after 2nd level) vocational track, 5 = special needs (for disabled) HS degree, 6 = HS drop-out (after 2nd level) technical track, 7 = HS graduate in general or artistic track, 8 = HS graduate in vocational track, 9 = HS graduate in technical track, 10 = tertiary vocational education, 11 = professional bachelor's degree, 12 = academic bachelor's degree, 13 = master's degree. (*) This is based on the variable "number of months of employment experience in the last 2 years. As mentioned in Section 3.2, those with "no employment experience" also include workers who were employed since graduation. However, because in this table reported groups almost always are born abroad, these workers most likely did not have any employment experience in Belgium.

When programme participation is limited to existing capacity, the simple rule based on a tree of depth 3 performs well in terms of employment gains compared to the black-box approach with sequential optimization. It attains nearly 80% (15.0/19.3=.78) of the employment gains that the latter approach achieves for switchers relative to the observed allocation. From Table 6.2 we can deduce that essentially recent migrants older than 28.2 are assigned to training according to this simple rule. Unemployed originating from an Eastern EU country are assigned to vocational training: those with less than 10 months of employment experience in Belgium[17] in the last two years to SVT and the other ones to LVT. This aligns with our findings in Sections 5.2 and 5.3, where we, respectively, found that the GATEs for SVT where highest for unemployed originating from an Eastern EU country and that programmes were relatively more effective for those with relatively little recent employment experience in Belgium.

Again, in line with our finding in Section 5.2 that OT is particularly effective for unemployed with no knowledge of Dutch, the allocation rule assigns all foreigners, except those originating from an Eastern EU country, to OT. Nevertheless, the rule that emerges from a decision tree of depth 3 does not perform as well with respect to the other component of the objective function. It results in only slightly more than 50% (-4.2/-8.0=.52) of the reduction of unemployment duration that switchers attain in the best performing black-box scenario.

The decision tree of depth 4 improves on that of depth 3 essentially by raising the overall performance, in both employment and unemployment duration, to about 80% of the best performing black-box rule: for employment the relative loss slightly increases from 22% to 24%, while for unemployment it falls substantially from 48% to 16%. Nevertheless, the marginal gain of moving from a decision tree of depth 3 to one of depth 4 improves the overall objective by only 11% (from 70% to 78% relative to the best black-box rule), while it complicates the rule quite substantially as can be deduced from Table 6.2.[18] Again, as a general rule mostly recent immigrants are assigned to the programmes. The age threshold is set at a higher level (34.9 instead of 28.2). Now also younger unemployed are assigned to LVT in case of poor proficiency in Dutch if they are born in a non-Southern EU country (excluding Belgium) and if they worked at least one month in the past two years in Belgium. Among the older group, the

[17] Recall that in Belgium individuals can qualify for UI based on foreign employment (see Section 2).

[18] A potential explanation for this complication is that the programme effectiveness in term of time spent in unemployment and employment may depend on different factors. Simplifying the objective by assigning all weight to either employment or unemployment may avoid this, but investigating this is beyond the scope of this research.

educational attainment matters in the assignment rule, but it is not intuitively clear why one group should be assigned to LVT, while the other group rather to SVT or OT. The group assigned to LVT is born in an Eastern EU country or in the rest of the world and started unemployment in December 2014, the first month that entries in unemployment are considered in the analysis. Those to be assigned to SVT are born in a non-Southern EU country or in the rest of the world and have not been employed in Belgium within the last two years. Those assigned to OT are born in the other foreign countries as well as in Belgium, but only to the extent that they have a poor proficiency in Dutch.

Finally, while the policy tree of depth 3 leads to a simple rule that yields about 70% of the gains of the best black-box rule, it may still be contested on different grounds. For instance, the use of country of birth and knowledge of Dutch may not be legally or politically acceptable. If so, this does not compromise the optimal policy approach. Quite on the contrary, it has made the rule more transparent, which is key in making its use socially acceptable or in defining restrictions that could be fed into the algorithm to determine a simple rule that is ethically and politically acceptable.

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 of participating in a future training programme within a preceding unemployment spell. 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: 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.

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. Moreover, we do not find evidence of effect heterogeneity in any dimension, in particular not in country of birth, Dutch proficiency, or employment experience within the last two years, which are dimensions that matter most in the main analysis.

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 been 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 issues. 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.

7.4 Comparison to propensity score matching

We also compared the results of the MCF for the ATE with conventional matching estimates on the propensity-score-matching-with-bias-adjustment, an estimator suggested by Lechner, Miquel, and Wunsch (2011) that showed were good performance in the extensive Empirical Monte Carlo study of Huber, Lechner, and Wunsch (2013). While the detailed the results are shown in Appendix C.4, overall point estimates as well as standard errors lead to very similar conclusions as for the MCF. Of course, such a comparison could only be done for the ATE (or ATEs for some large subgroups), as such estimators are not adapted for the estimation of more disaggregated effects.

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-in 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. 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 used our estimates of individual programme effects to evaluate the existing allocation of the Flemish PES to the considered training programmes. We found that changing the assignment rules could increase the time that reallocated individuals spend in employment by about 20%. This is a substantial gain and illustrates the social value of using CML methods in the allocation of unemployed to active labour market policies. Moreover, we have shown how this optimal allocation can attain about 70% of the overall gain in terms of time spent in employment and unemployment with a policy tree of depth 3, while a policy tree of depth 4 results in only a modest gain relative to the increased complexity. One advantage of simple rules is that case-workers are more likely implement it, because they can easily understand it. Another advantage is that this makes the rule more transparent and, hence, more ethically or societally acceptable (Whittlestone, Nyrup, Alexandrova, Dihal, and Cave, 2019; Reddy, Allan, Coghlan, and Cooper, 2020). However, in this application the simple rule requires discrimination on grounds – country of birth and knowledge of Dutch – that might not be legally or politically acceptable. If so, this does not compromise the approach, but merely defines restrictions that could be fed into the algorithm that determines the simple rule.

Future work could address many open issues, such as extending the data base with (i) programme costs, so that the derivation of optimal policy rules can be based on the net social value of programmes, and with (ii) additional control variables, so 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 that it can address issues related to dynamic assignment is likely to lead to additional insights about policy effectiveness and assignment optimality.

References

Altonji, J. G., E. E. Todd, and C. R. Taber (2005): Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, *Journal of Political Economy*, 113(1), pp. 151-184.

Arthur, D., and S. Vassilvitskii (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 (2019): 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.

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): Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence, *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.

Lechner, M. and J. Smith (2007): What is the value added by case workers?, *Labour Economics*, 14, 135-151.

Piton, C., and F. Rycx (2020): The heterogeneous employment outcomes of first- and second-generation immigrants in Belgium, National Bank of Belgium discussion paper N°381, Brussels: National Bank of Belgium.

Reddy, S., S. Allan, S. Coghlan, and P. Cooper (2020): A governance model for the application of AI in health care, *Journal of the American Medical Informatics Association*, 27(3), 2020, 491–497.

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.

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.

Whittlestone, J., R. Nyrup, A. Alexandrova, K. Dihal, and S. Cave (2019): Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research, London: Nuffield Foundation.

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

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

Bart Cockx[$], Michael Lechner[$$], Joost Bollens[$$$*]

[$] Department of Economics, Ghent University
[$$] Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen
[$$$] Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding (VDAB)

This version: September2020

Date this version has been printed: ***29 September 2020***

Appendix for Online Publication

Addresses for correspondence:
Bart Cockx, Ghent University, Department of Economics, Sint-Pietersplein 6, 9000 Gent, Belgium, bart.cockx@ugent.be, sites.google.com/site/bartcockxsite.
Michael Lechner, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstr. 14, 9000 St. Gallen Switzerland, Michael.Lechner@unisg.ch, www.michael-lechner.eu, www.researchgate.net/profile/Michael_Lechner.
Joost Bollens, VDAB, Keizerslaan 11, 1000 Brussels, Belgium, joost.bollens@vdab.be.

---

[*] Bart Cockx is also affiliated with IZA Bonn, CESIfo, Munich, IRES, Université catholique de Louvain and ROA, Maastricht University. 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). We thank Michael Knaus, Anthony Strittmatter and Gabriel Okasa for carefully reading and commenting on a previous draft of the paper. Previous versions of the paper were presented at research seminars of the LMU University of Munich, the University of Milan, Competence Centre of Microeconomic Evaluation of the Joint Research Centre of the European Commission, and at the IMT in Lucca, as well as in workshops and conferences in Copenhagen, St Gallen, Engelberg, and Basel. We thank participants, in particular Bas Van der Klaauw, Claudio Lucifora, Regina Riphahn, Claus Schnabel, and Andreas Steinmayr, for helpful comments and suggestions. We thank the VDAB for providing us access to the data. The VDAB stocks the data and has engaged to make the data available to researchers who wish to replicate the current research. The usual disclaimer applies.

Appendix A: Data

A.1 Descriptive statistics by programme participation

Table A.1: Means and standardized differences[1] 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-benefit 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 benefit sanction before current unemployment spell (cat_14)	.06	.07	.04	.03
		(6.63)	(9.15)	(12.71)
Ever have had a sickness benefit 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 counselling before current spell (vroeger_t)	.06	.08 (5.04)	.05 (4.38)	.06 (.31)
Having participated in an 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: High school (HS) drop-out (after 1st level) (so1)	.14	.18 (10.29)	.04 (34.75)	.10 (13.63)

Table A.1, continued

Educ. Attainment: part-time HS degree, vocational track (dbso)	.04	.06 (9.51)	.03 (2.97)	.04 (1.73)
Educ. Attainment: HS drop-out (after 2nd level HS), general or artistic tracks (akso2)	.02	.02 (1.72)	.02 (.84)	.03 (4.66)
Educ. Attainment: HS drop-out (after 2nd level HS) , vocational track (bso2)	.09	.13 (13.53)	.06 (9.47)	.08 (2.65)
Educ. Attainment: Special needs (for disabled) HS degree (buso)	.02	.04 (11.29)	.01 (5.18)	.02 (1.69)
Educ. Attainment: HS drop-out (after 2nd level HS) technical track (tso2)	.04	.05 (4.11)	.04 (.02)	.05 (8.08)
Educ. Attainment: HS graduate, general or artistic track (akso3)	.08	.08 (1.39)	.13 (16.86)	.11 (10.42)
Educ. Attainment: HS graduate, vocational track (bso3)	.21	.26 (13.3)	.21 (1.41)	.22 (4.08)
Educ. Attainment: HS graduate, technical track (tso3)	.11	.11 (.27)	.22 (29.68)	.17 (17.82)
Educ. Attainment: Tertiary vocational education (hbo)	.02	.01 (8.48)	.02 (4.01)	.02 (.52)
Educ. Attainment: professional bachelor's degree (pba)	.13	.04 (31.38)	.14 (4.19)	.11 (4.89)
Educ. Attainment: academic bachelor's degree (aba)	.02	0.00 (11.01)	.01 (2.08)	.01 (1.02)
Educ. Attainment: academic master's degree (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	1,220	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 $j \in \{SVT, LVT, OT\}$.

[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.

A.2 Analysis of selection between training programmes

Table A.2: Means and standardized differences (SD)[1] of conditioning variables for which the maximum SD of LVT and OT relative to SVT is larger than 20%

	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
Conditioning variables			
Gender (female =1)(Woman)	.31	.40	.47
		(18.99)	(32.48)
Knowledge of French (frans)	.54	.72	.67
		(39.04)	(27.04)
Knowledge of English (engels)	.66	.82	.79
		(37.30)	(28.23)
Knowledge of German (duits)	.19	.34	.32
		(32.65)	(29.59)
Proficiency in Dutch² (Lang_dutch)	2.51	2.72	2.64
		(34.50)	(20.34)
Ever been on welfare before current unemployment spell (cat_5)	.14	.04	.08
		(33.35)	(18.08)
Drivers license car (rijbew_B)	.63	.76	.71
		(27.27)	(16.72)
Educ. Attainment: High school (HS) drop-out	.18	.04	.1
(after 1st level) (so1)		(44.6)	(23.86)
Educ. Attainment: HS drop-out (after 2nd level HS) , vocational track (bso2)	.13	.06	.08
		(22.85)	(16.15)
Educ. Attainment: HS graduate,	.11	.22	.17
technical track (tso3)		(29.94)	(18.09)
Educ. Attainment: professional bachelor's degree (pba)	.04	.14	.11
		(35.36)	(26.70)
District of residence: Tongeren (Tongeren)	.03	.03	.09
		(0.80)	(24.80)
Country of birth: Belgium (belg)	.66	.77	.71
		(25.92)	(10.91)
Country of birth: Rest of the world (row)	.21	.10	.14
		(28.8)	(16.35)
Preferred profession 2³ (prof2)	.04	.20	.20
		(54.05)	(53.94)
Preferred profession 29³ (prof29)	.07	.03	.03
		(20.74)	(18.83)
Preferred profession 31³ (prof31)	.12	.06	.06
		(18.88)	(21.17)
Preferred profession 32³ (prof32)	.01	.06	.06
		(26.99)	(25.94)

Table A.2, continued

Preferred profession 33[3] (prof33)	.01	.04	.04
		(20.42)	(18.47)
Preferred profession 43[3] (prof43)	.08	.03	.03
		(25.09)	(20.86)
Number of observations	1,305	1,220	1,115

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

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

[3] 2 = General clerk; 29 = Hall staff; 31 = Construction workers and technicians; 32 = Specialised administrative staff; 33 = Computer and ICT staff; 43 = Personal service providers, Cleaning and maintenance personnel.

Table A.2 reports the means and SD of conditioning variables for which the maximum SD of LVT and OT relative to SVT is larger than 20%. By considering only variables for which these SD are larger than 20%, we can study in which direction there is imbalance of participants in SVT relative to the other two programs. We first summarize for each set of variables the direction of the selection bias of SVT relative to LVT and then list the associated SD for each variable (we only retain variables for which the corresponding SD is larger than 20%) and between parenthesis first the mean of participants in SVT versus (vs) the mean of participants in LVT. Next, after a brief conclusion what we can learn from this analysis, a brief comparison is made between the participants in LVT and OT.

1) With respect to education, we find that participants in SVT have a much lower level of educational attainment than those in LVT: (i) high school drop-out (level 1): SD=44.6 (.18 vs .04); (ii) HS drop-out – vocational track (level 2): SD=22.85 (.13 vs .06); (iii) HS graduate technical track: SD=29.94 (.11 vs .22); (iv) professional bachelor's: SD 35.36 (.04 vs .14).
2) There are fewer Belgians and more individuals originating from the rest of the world participating in SVT than among those participating in LVT: (i) Belgian: SD=25.92 (.66 vs .77); (ii) ROW: SD=28.8 (.21 vs .10).
3) The Dutch proficiency is lower for participants in SVT than in LVT: SD=34.5 (2.51 vs 2.72).

4) The proficiency in (useful) foreign languages is lower for participants in SVT than in LVT: (i) French: SD=39.04 (.54 vs .72); English: SD=37.3 (.66 vs .82); German: SD=32.65 (.19 vs .34).
5) A higher fraction of participants in SVT has ever been on welfare than those in LVT (cat_5): SD=33.35 (.14 vs .04).
6) Fewer participants in SVT have a driver's licence for a car than those in LVT: SD=27.27 (.63 vs .76).
7) Some of previously occupied professions for which the labour market is typically tight are less represented among participants in SVT than in LVT: (i) general clerk (prof2): SD=54.05 (.04 vs .20); Specialized administrative staff (prof32): SD=26.99 (.01 vs .06); Computer and ICT staff (prof33): SD=20.42 (.01 vs .04).
8) Other professions, for which applicants typically exceed candidates, are more represented among participants in SVT than in LVT: (i) hall staff (prof29): SD=20.74 (.07 vs .03); (ii) Personal service providers, Cleaning and maintenance personnel (prof43): SD=25.09 (.08 vs .03).

Conclusion: Participants in SVT have generally characteristics that are less valuable on the labour market than participants in LVT. This means that they are a *negative* selection in terms of observables relative to participants in LVT.

We can also deduce from Table A.2 that participants in OT have a very similar (or maybe slightly worse) profile as those in LVT, which is, hence, systematically better than participants in SVT.

Appendix B: Additional details of estimation

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 9 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 variables 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 delete 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 VIMs 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.

B.4: Algorithm used for computing policy trees

We modify Algorithm 2 of Zhou, Athey, and Wager (2019) in three ways. First, instead of using one global approximation level for computing possible splitting rules (*A* in their notation), we use a finer grid at higher levels of the tree (1st level, i.e. at the bottom of the tree: *A*, 2nd level: *A/2*, 3rd level, *A/4*, 4th level, i.e. at the top of the tree *A/8*). Second, we enforce constraints with respect to maximum individual programme shares as well as overall programme shares. Third, we allow explicitly for unordered categorical variables.

The algorithm implemented can be written in three parts. The augmented baseline algorithm of Zhou, Athey, and Wager (2019) is denoted by Algorithm 1. Algorithm 2 is the module enforcing constraints, while Algorithm 3 deals with categorical variables. While in the presentation of Algorithm 1, we close follow Zhou, Athey, and Wager (2019), we also keep the description at a stylised level and abstract from many implementational details.

A few remarks are in order about the elements that differ from the original proposal that essentially all address some approximation that reduce the computational costs of finding ‘good’ trees that are close to the optimal ones. With respect to algorithm 1 reducing A at higher levels leads to additional precision at reasonable additional computational costs. This precision gain is more relevant when fewer splitting points are available as in the higher levels.

Second, the idea of enforcing the constraints is that whenever a subtree violates the constraints, treatments are switched to the next best alternative for a leave until there is no longer any violation. In this process, the order matters. Here we order the leaves according to their within-leaf variance of the potential outcomes. This gives a computationally convenient approximation whether there is large loss with respect to the outcome variables when treatments are moved away from the best option. Therefore, switching starts with the leaf with the lowest variance and stops once the constraints are fulfilled for the (sub-) tree. Clearly, this sequential process may not identify the optimal solutions, which is however very computationally intensive to obtain in case of such types of constraints.

Finally, to deal with categorical variables, we follow a modified text book treatment (i.e. Hastie et al., 2009). The difference is that sorting is not with respect to the potential outcomes but with respect to the within-value-of-covariate variance as an approximate measure for possible gains and losses from reallocation.

Algorithm 1: Tree search

Input: $\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}_{i=1}^N$, depth *L,* approximation parameter *A,* max category value *V,* restrictions *R, total number of variables p*
If *L* = 1
Return $\left(\max_{j\in\{0,\dots,3\}} \sum_{i=1}^N \hat{Y}_i^j, \quad \arg\max_{j\in\{0,\dots,3\}} \sum_{i=1}^N \hat{Y}_i^j\right)$
Else
Initialise *reward* = -∞, *tree* = Ø, reduce approximation parameter: *A* = *A/2*
For *m* = 1,2,…,*p*
If m^{th}-coordinate of *X* is an ordered variable
Sort X_i according to the m^{th}-coordinate
If m^{th}-coordinate of *X* has less than *V* different values
k = vector of all different values of m^{th} -coordinate of *X*
Else
k = vector of every A^{th} value of m^{th}-coordinate of X_i (drop all elements equal to the previous value)
Endif
Else
(X_i, k) = CATEGORICAL VARIABLES($\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}_{i=1}^N$, *m*)
Endif
For *ii* = 1, 2, … dim(k)-1
If m^{th} -coordinate is ordered with more than *R* categories
$\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{left}$ for *i* where m^{th} -coordinate of *X* is smaller or equal than the i^{th} element of *k*
$\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{right}$ for *i* where m^{th} -coordinate of *X* is larger than the ii^{th} element of *k*
Else
$\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{left}$ for *i* where m^{th} -coordinate of *X* has values of *1st* to ii^{th} elements of *k*
$\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{right}$ for *i* where m^{th} -coordinate of *X* has values of ii^{th}+1 to $dim(k)^{th}$ elements of *k*
Endif
(reward_left, tree_left) = TREE SEARCH $\left(\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{left}, L-1, A, V, R\right)$
(reward_right, tree_right) = TREE SEARCH $\left(\left\{X_i, \hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{right}, L-1, A, V, R\right)$
If reward_right + reward_left > reward
(reward, tree) = IMPOSE RESTRICTIONS (tree_left, tree_right, return_left, return_right, $\left(\left\{\hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{left}\right)$, $\left(\left\{\hat{Y}_i^0, \hat{Y}_i^1, \hat{Y}_i^2, \hat{Y}_i^3\right\}^{right}\right)$, reward, tree, *R*)
Endif
Endfor
Endfor
Return (reward, tree)
Endif

Note: *V* is set to the starting value of *A*.

Algorithm 2: Impose restrictions

Input: tree_left, tree_right, return_left, return_right, $\left\{\hat{Y}_i^0,\hat{Y}_i^1,\hat{Y}_i^2,\hat{Y}_i^3\right\}^{left}, \left\{\hat{Y}_i^0,\hat{Y}_i^1,\hat{Y}_i^2,\hat{Y}_i^3\right\}^{right}$, reward_old, tree_old, *R*

return = return_left + return_right

tree = tree_left & tree_right

If tree fulfils constraints *R*

Return (reward, tree)

Else

For each leaf of tree: compute variance of the 4 within-leaf averages of $\left\{\hat{Y}_i^0,\hat{Y}_i^1,\hat{Y}_i^2,\hat{Y}_i^3\right\}$

Sort leaves from smallest to largest variance

For *l* = # of leaves

If treatment share of treatment assigned to leaf *l* violates *R*

For *d* = 1, 2, 3, 4

Reassign next best treatment to leaf

tree = Recompute tree with changed treatment assignment

If tree fulfils *R*

Breakout of loop

Endif

Endfor

If tree fulfils *R*

Breakout of loop

Endif

Endfor

Reward_new = reward of tree

If reward_new > reward_old

Return (reward_new, tree)

Else

Return (reward_old, tree_old)

Endif

Algorithm 3: Categorical variables (in tree search)

Input: $\left\{X_i,\hat{Y}_i^0,\hat{Y}_i^1,\hat{Y}_i^2,\hat{Y}_i^3\right\}_{i=1}^{N}$, *m*

$Var_i = Var\left(\hat{Y}_i^0,\hat{Y}_i^1,\hat{Y}_i^2,\hat{Y}_i^3\right)$

Var_mean_k = vector of average values of Var_i *for* all different values of m^{th}-coordinate of *X*

k = sorted different values of m^{th}-coordinate of *X* according to values of Var_mean_k

Sort X in the same way as k

Return (X_i, *k*)

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.

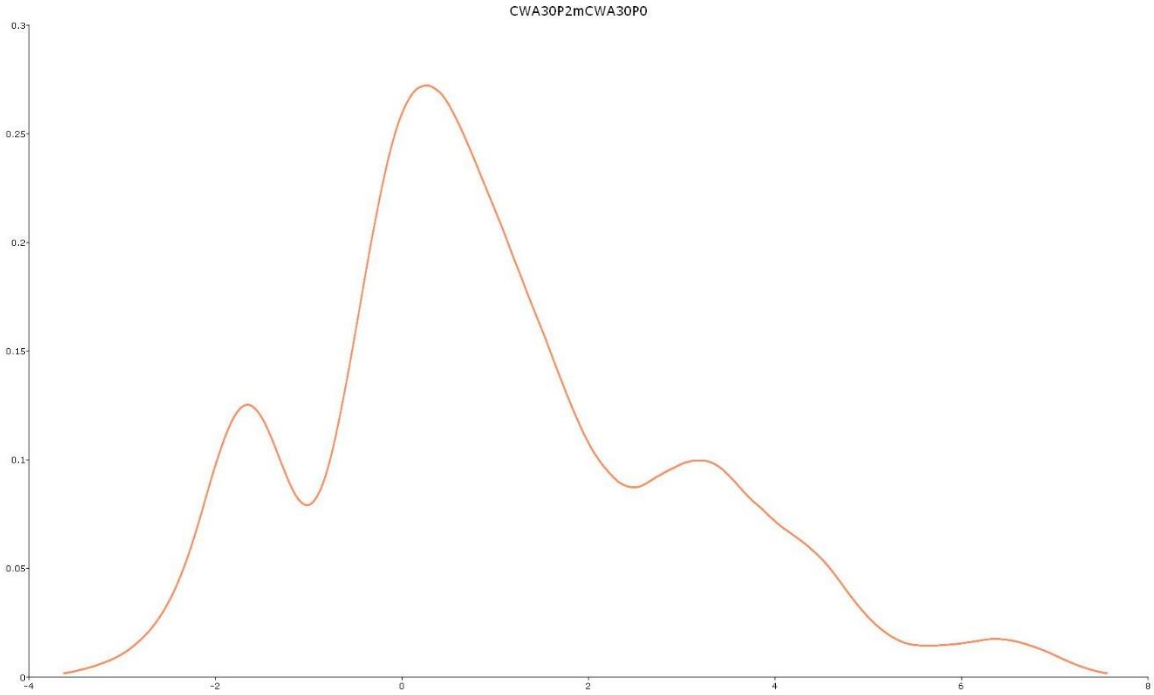
Table C.1: Comparison of the effects for the different programmes on cumulative months in employment, unemployment and out of the labour force for different populations by participation status

	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.

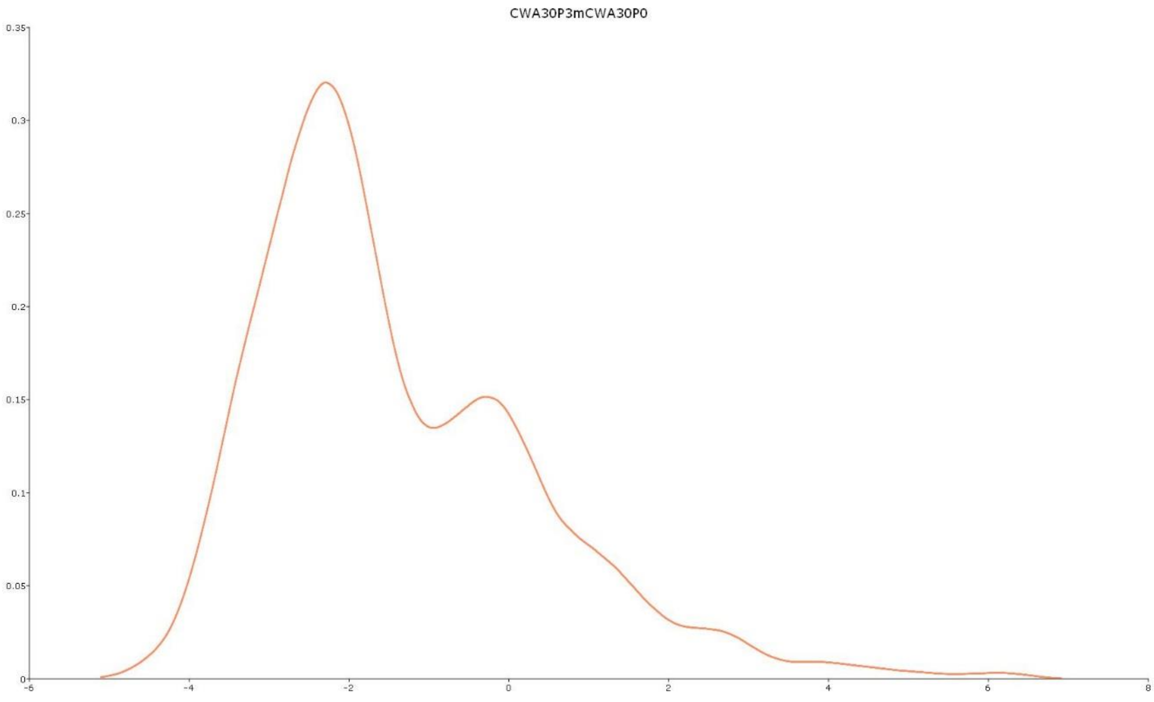
C.2 Distribution of estimated IATEs

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



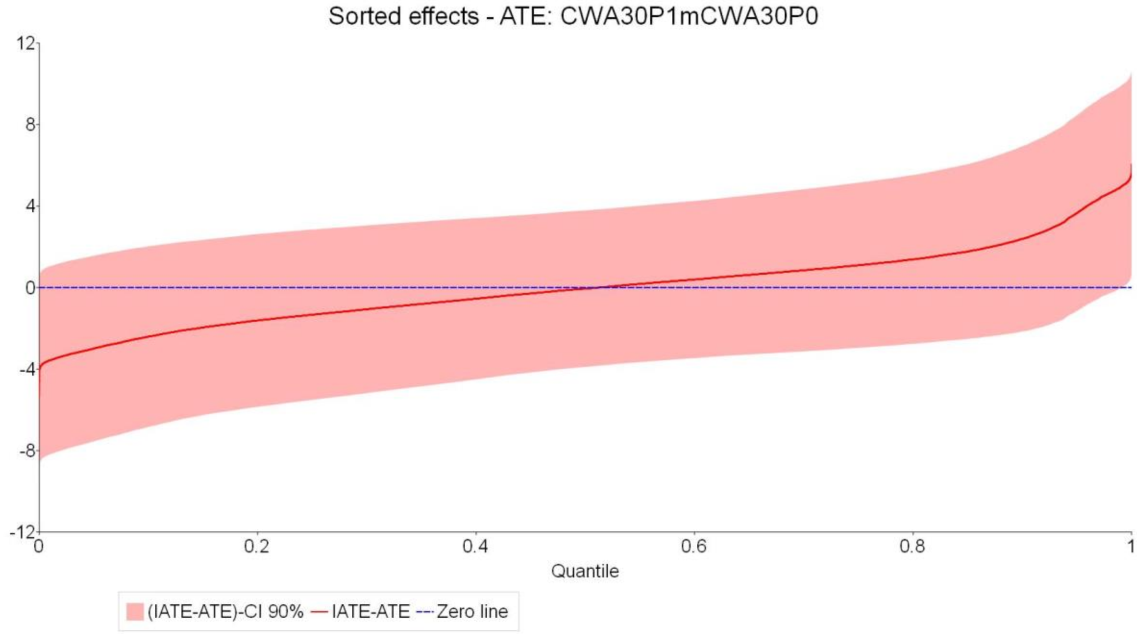
Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.

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



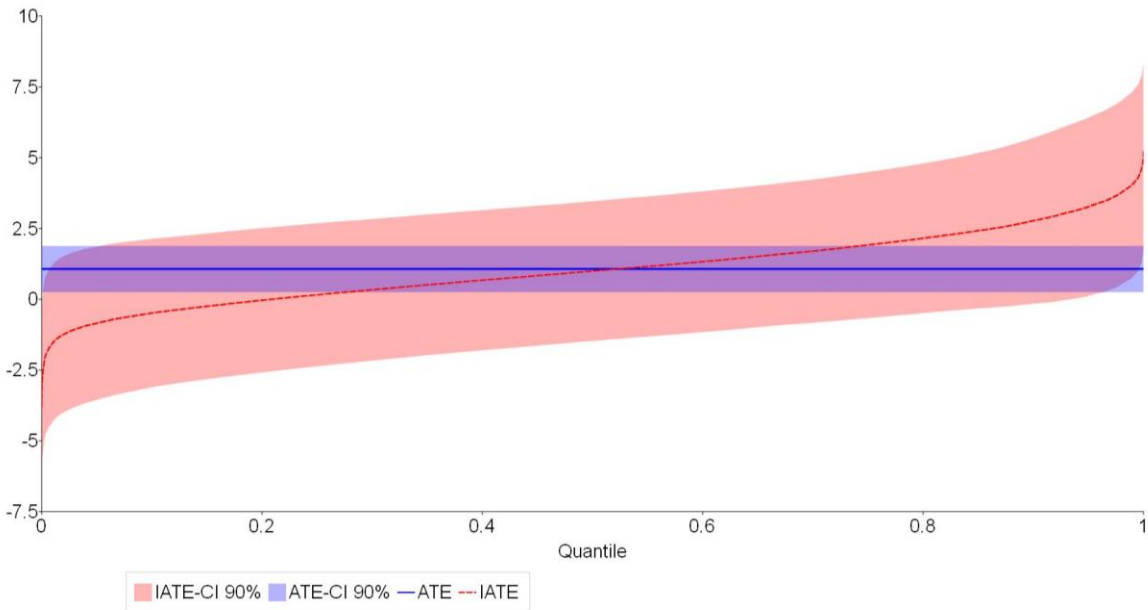
Note: Kernel smooth with Epanechnikov Kernel and Silverman (normality) bandwidth.

*Figure C.4: Overall heterogeneity: sorted effects **minus ATE** of SVT relative to NOP – Employment 30 month after the programme start*



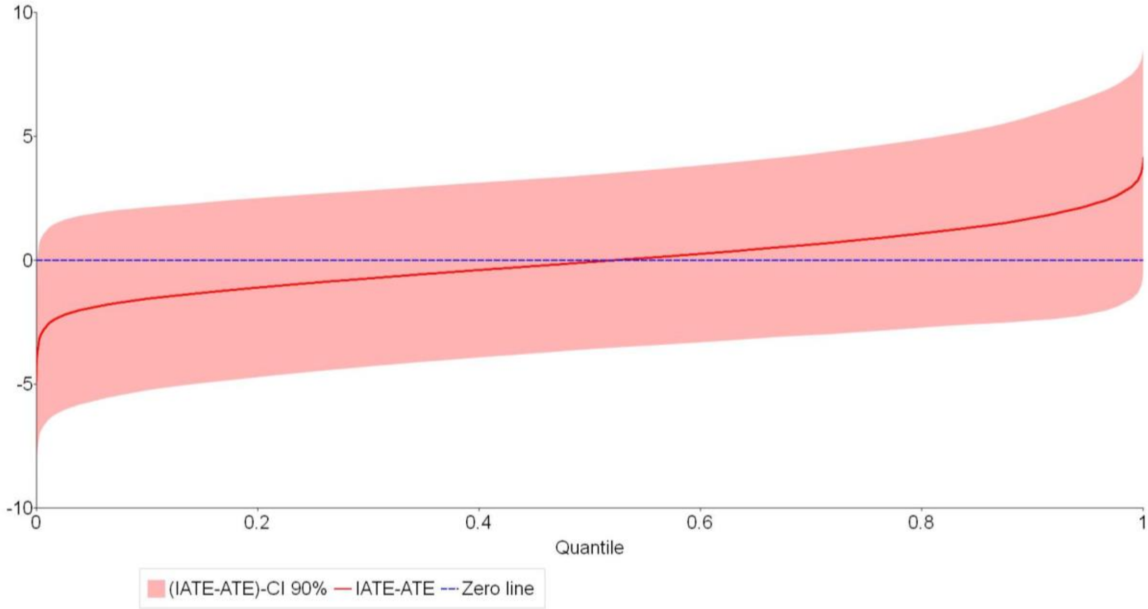
Note: (IATEs-ATE) are sorted according their size. 90%-confidence interval based on estimated standard errors and normal distribution. Standard errors are smoothed by Nadaraya-Watson regression (Epanechnikov kernel with Silverman bandwidth).

Figure C.5: Overall heterogeneity: sorted effects of LVT relative to NOP – Employment 30 month after the programme start



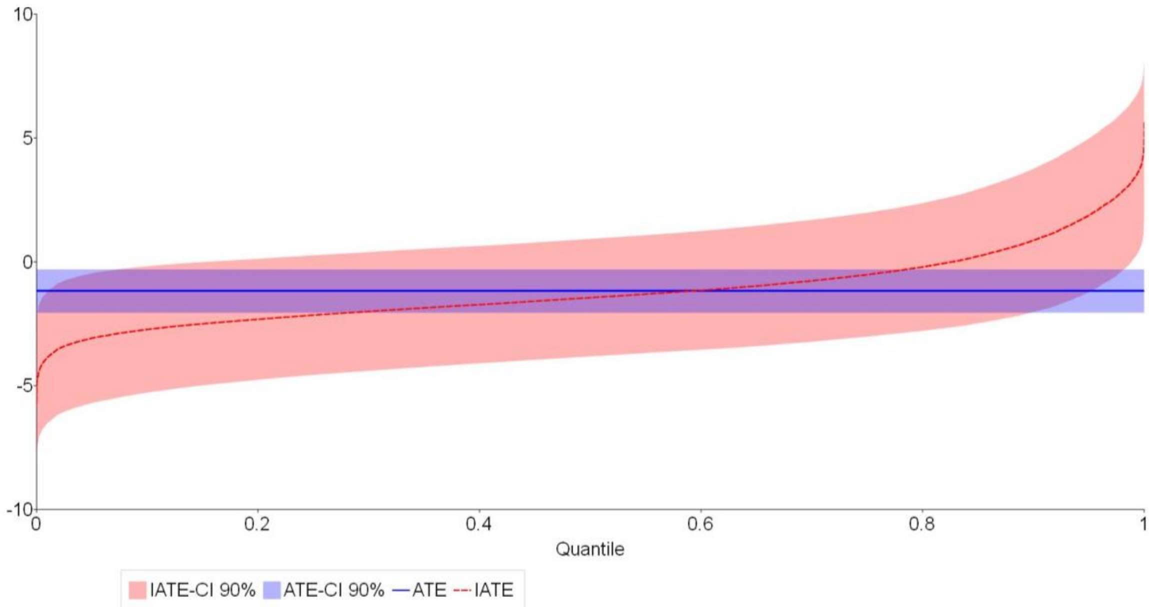
Note: IATEs are sorted according their size. 90%-confidence interval based on estimated standard errors and normal distribution. Standard errors are smoothed by Nadaraya-Watson regression (Epanechnikov kernel with Silverman bandwidth).

*Figure C.5: Overall heterogeneity: sorted effects **minus ATE** of LVT relative to NOP – Employment 30 month after the programme start*



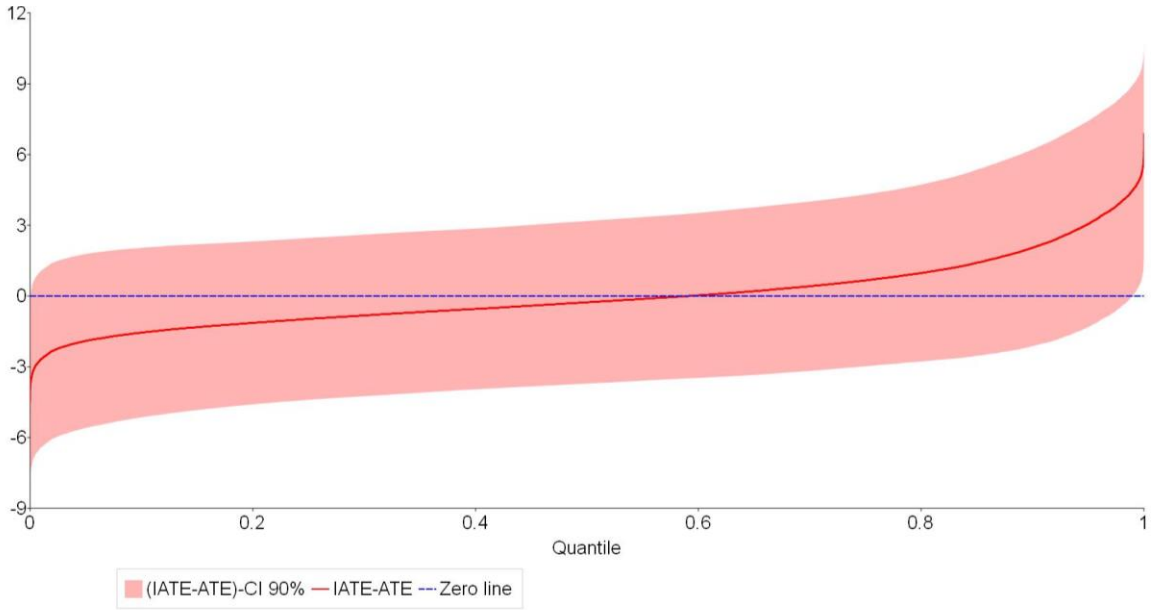
Note: (IATEs-ATE) are sorted according their size. 90%-confidence interval based on estimated standard errors and normal distribution. Standard errors are smoothed by Nadaraya-Watson regression (Epanechnikov kernel with Silverman bandwidth).

Figure C.6: Overall heterogeneity: sorted effects of OT 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).

*Figure C.7: Overall heterogeneity: sorted effects **minus ATE** of OT relative to NOP – Employment 30 month after the programme start*



Note: (IATEs-ATE) 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).

C.3 Additional policy simulations

Table C.2 contains the full set of policy simulations. The results of Table 6.1 in the main text are also contained in this table. We focus our discussion on the scenario's that are not discussed in the main text. We first consider the Black-Box approaches. We did not consider the following three constrained scenarios' in which in case priority is given to unemployed with (i) worst NOP outcomes; (ii) poor proficiency in Dutch; and (iii) poor proficiency in Dutch who are born abroad, which is a proxy for "recent immigrants". We contrast these scenario's each time to the close to "optimal" scenario, i.e. *constrained*, *sequential optimization* ("scenario opt"), and to the priority rule *constrained*, *preference to lots of past UE* ("scenario 0"). When we compare for scenario (i) the gains for reallocated individuals in terms of time spent in employment and unemployment increases by about 50% relative to scenario 0, but it still performs only half as well as scenario opt. The improvement relative to scenario 0 was expected, because, as clearly observed in the last line of Table 5.3, we see that programme effectiveness increases with the counterfactual NOP outcome, while it decreases with the time spent in UE. Nevertheless, this simple priority rule is still much less effective than the optimal rule. Scenario (ii) illustrates that giving priority to a dimension which increases effectiveness according to Table 5.3, namely giving priority those who have a bad knowledge of Dutch, is no guarantee for more successful outcomes, because we find that it performs even worse than scenario 0: E.g., the gain for switchers in terms of time spent in employment is only 4.4% compared to 6.6% for scenario 0. This is presumably related to the fact that a nonignorable fraction of unemployed with poor Dutch proficiency are Belgians, for whom the programme effectiveness is much lower (see Table 5.3). However, when bad language proficiency is combined with being foreign born, i.e. when we give priority to recent migrants (scenario (iii)), then performance improves and is for employment nearly as good as scenario (i), but still much below scenario opt. This also suggests that a simple rule that gives priority to recent immigrants would reap less than half of the benefits of those obtained with the optimal rule.

Table C.2: Overall effects of simulated hypothetical programme allocations

	Share of different programmes in %			Cumulative # of months 30 months after programme start			Gain for switchers in %	
	SVT	LVT	OT	Emp	OLF	UE	Emp	UE
Observed	2.1	2.0	1.8	16.1	3.0	10.9	-	-
Random	2.0	1.9	1.9	16.1	3.0	10.9	0.9	-0.1
Black-Box – no constraint	97.3	0.2	0.2	19.4	1.6	8.9	21.9	-18.5
Black-Box – no constraint, only significant	58.1	1.6	0.5	18.8	1.8	9.4	33.4	-23.0
Black-Box – constrained, preference to large gains	2.1	2.0	1.8	16.4	2.8	10.8	13.6	-6.3
Black-Box – constrained, sequential optimization	2.1	2.0	1.8	16.4	2.9	10.8	19.3	-8.0
Black-Box – constrained, preference to worst NOP outcomes	2.1	2.0	1.8	16.3	2.9	10.8	9.9	-3.6
Black-Box – constrained, preference to lack of language skills*)	2.1	2.0	1.8	16.2	3.0	10.9	4.4	-2.2
Black-Box – constrained, preference to "recent immigrants**)	2.2	2.1	1.8	16.3	2.9	10.9	9.3	-2.4
Black-Box – constrained, preference to lots of past UE***)	2.1	2.0	1.8	16.2	2.9	10.9	6.6	-2.5
Simple – Policy tree 3 level, overall number of participants same	5.9	0	0	16.5	2.9	10.6	24.2	-19.4
Simple – Policy tree 3 level, constrained	2.0	2.0	1.8	16.3	2.9	10.8	15.0	-4.2
Simple – Policy tree 4 level, constrained	2.1	2.0	1.8	16.4	10.8	2.8	14.6	-6.7

Note: Emp: Employed; OFL: Out-of-labour-force; UE: Unemployed
Allocations minimize unemployment and maximise employment (equally weighted). If programmes capacity becomes a constraint, preference is given to *) those UE who have no or limited proficiency in Dutch, **) "Recent migrants" are proxied by individuals born abroad and having no or limited knowledge of Dutch. ***) the highest number of months in unemployment over the last 10 years. *Sequential optimisation* means that starting with the observed allocation programmes states are pairwise switched if overall outcome is improved and budget constraint is kept.

Table C.2 contains policy a tree scenario of depth 3 that is not reported in the main text. It considers a scenario in which the programme capacity constraint is not fixed by each individual programme, but rather over all programmes together. In this case all available training slots would be assigned to SVT, which is according to expectations, as SVT clearly dominated the two other programmes in terms of effectiveness. The average performance of this scenario lies between the unconstrained and the constrained one that was reported in the main text.

C.4 Matching results

In Table C.3 we report the estimates of the ATE on the cumulative number of months in employment 30 months after the programme start, i.e. our benchmark, based on the conventional propensity score radius matching method with bias adjustment as suggested by Lechner, Miquel and Wunsch (2011). We only report the programme effects relative to No ALMP participation (NOP). The same set of matching variables are chosen as in the MCF, but no polynomials or interactions are considered. The categorical variables are broken up into dummy variables and some of these dummies merged in case of too small groups. The comparisons are based on slightly different common supports. The standard errors are analytical ones. These are known to be slightly conservative.

The estimated ATE are not very different from the ones obtained by MCF reported in Table 5.1 in the main text. The point estimated for the ATE of OT is nearly identical: -1.3 versus -1.4 by MCF. For LVT the point estimate obtained by the propensity score method is slightly smaller than the one obtained by MCF, but well within one standard error of the MCF estimate: 0.8 versus 1.0. Finally, the differences between the point estimates of the ATE of SVT is somewhat larger than for the other programmes, but the difference remains well within two standard errors of the MCF estimate: 2.6 versus 3.4 for the MCF estimate.

Table C.3: Estimates of ATE using the bias corrected propensity score estimator. Reference is No ALMP participation (NOP).

	Short vocational training (SVT)	Long vocational training (LVT)	Orientation training (OT)
ATE (relative to NOP)	2.61 (0.51)***	0.83 (0.49)*	-1.34 (0.48)**

Note: Outcomes measured in months. All effects are population averages (ATE) relative to No ALMP Participation (NOP). Analytical 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.