Comparing methods to evaluate the effects of job search assistance

Paul Muller\textsuperscript{a}\textsuperscript{*} Arjan Heyma\textsuperscript{**} Bas van der Klaauw\textsuperscript{a}.

December 23, 2014

Abstract

In March 2010, budgetary problems at the Dutch unemployment insurance administration caused a sharp drop in the availability of job search assistance programs. We exploit this discontinuity in the participation rate to (non-parametrically) estimate the effects of the program on job finding. Next, by using a large and rich administrative data set provided by the UI administration, we estimate the same program effects using non-experimental methods (without using the exogenous variation due to the policy discontinuity). We compare the results of the two approaches and discuss how the difference relates to the identification assumptions required by the non-experimental methods. All methods find a negative effect on job finding during the first six months after the program starts, but only the experimental estimator and the timing-of-events model show that the medium-run effect is zero.

Keywords: Policy evaluation, Job search assistance, Unemployment duration
JEL-code: J64, C14, C31

1 Introduction

In 2002 the Dutch market for job search assistance programs was privatized, implying that the unemployment insurance (UI) administration buys the services of private companies to guide benefits recipients to employment. Due to the economic crisis the

\textsuperscript{a} Corresponding author. VU University Amsterdam, Department of Economics, De Boelelaan 1105, 1081 HV Amsterdam. Email: p.muller@vu.nl. https://sites.google.com/site/researchpaulmuller/home. This paper has benefitted from comments by Gréory Jolivet, Jaap Abbring, Mike Elsby and seminar participants at the VU University, University of Edinburgh, BeNA (Berlin), EALE (2013) and ESPE (2014).
number of programs offered increased sharply in 2009 and early 2010. In March 2010 the government suddenly refused to extend the budget. As a result, the purchase of new programs was terminated within a period of two weeks. During the remainder of the year, hardly any programs started. In this paper we exploit this policy discontinuity to evaluate the effects of the programs on job finding. We then estimate the same effects using non-experimental methods and assess their performance using the experimental estimates as a benchmark.

The main challenge faced by studies evaluating activation programs is selective participation (Heckman et al. (1999), Abbring and Heckman (2007)). As shown in a meta-analysis by Card et al. (2010), over 50% of such studies use longitudinal data and compare a treatment group with a control group, where the control group is typically formed by matching on observable characteristics. Over one-third of the studies use duration models. Less than 10% of the studies use an experimental design. In his seminal study, LaLonde (1986) shows that non-experimental estimators produce results that do not concur with those from experimental evidence. In an extension Dehejia and Wahba (1999) test the performance of matching estimators, finding that those are closer to the experimental evidence. Smith and Todd (2005) evaluate the same program and show that these findings are not robust to different specifications and to the use of different samples and different sets of covariates. They refer to Heckman et al. (1997) who argue that matching estimators can only replicate experimental findings if three requirements are fulfilled. First, the same data source for treated and control group should be used (in particular the outcome variable should be measured in the same way). Second, treated and control individuals should be active in the same local labor market. And third, the data should contain a rich set of variables that affect both program participation and labor market outcomes. Smith and Todd (2005) state that each of these requirements is likely to be violated in the evaluations by LaLonde (1986) and Dehejia and Wahba (1999). Each of these papers evaluates an employment program from the 70’s. Data quality has improved since then, as have statistical methods.

We build on this literature by performing a similar comparison of methods, using a recent, rich administrative dataset. Our main contribution is twofold. First, since our experimental estimates are identified from a large scale policy discontinuity in 2010 in the Netherlands, the setting is particularly suitable for such a comparison. Our data fulfills the criteria mentioned by Smith and Todd (2005). The administrative dataset allows the use of high-quality information on a rich set of variables, including individual characteristics, pre-unemployment labor market variables, current spell characteristics and any assistance provided by the UI administration and
private providers. As the policy discontinuity was nationwide, we have a large sample. Since this is a recent setting, programs and labor market conditions are similar to those currently in many countries, making the comparison of methods particularly interesting. Second, for the non-experimental analysis we not only apply a range of matching estimators, but also estimate the timing-of-events model. So we can compare both approaches to our baseline estimates and assess which performs best. As such we provide an assessment of how non-experimental methods perform, given the current quality of data, and given recent active labor market programs and labor market conditions.

The policy discontinuity allows us to non-parametrically estimate how program participation affects job finding. The experimental variation in program provision is large. Within a month, the weekly number of new programs dropped from 1,300 to less than 80 and remained below 50 for the remainder of the year. This sharp discontinuity enhances non-parametric identification of the programs effect on job finding. We can estimate the treatment effect on the treated for a large sample of treated individuals, and are not restricted by small-scale or local variation. We compare the job finding rates of cohorts entering unemployment at different points in time, though relatively short after each other. Since they reach the discontinuity at different unemployment durations they are affected differentially. This identifies the effect of the programs. Seasonal differences in the labor market are controlled for using cohorts from the previous year. Our results show that after starting a program, the job finding rate is reduced significantly for several months. After half a year it increases, up to a zero difference in job finding after 12 to 18 months.

We compare these results to several matching estimators (propensity score matching, inverse probability weighting, regression adjustment and nearest neighbor matching). Results are very similar across the four matching estimators. All show a significant negative effect of program participation directly after the program has been started. Even though the negative effect decreases in magnitude over time, all estimators are still significantly negative after 18 months. Next, we estimate a timing-of-events model (Abbring and Van den Berg (2003)), which adds more structure by jointly modelling the hazard rate for employment and the hazard rate for program participation. Both hazard rates include an unobserved heterogeneity term which are allowed to be correlated. Using this model we estimate the effect of the program to reduce the job finding rate in the first six months, while it slightly increases the job finding rate at longer durations. Overall this leads to a negative effect on employment in the first 14 months, and a zero effect afterwards. So all methods find a negative effect in the first couple of months, while only the experi-
mental estimates and the timing-of-events model find that the effect at the medium run is zero. The fact that the matching estimator underestimates the effect suggests that negative selection on unobservable characteristics occurs.

The rest of the paper is structured as follows. We briefly discuss the literature on the evaluation of active labor market programs in section 2, and describe the institutional setting and the budgetary problems which led to the policy discontinuity in Section 3. An overview of the data is provided in section 4. In section 5 we define the treatment effects and discuss how they are identified from the policy discontinuity. In section 6 we present the experimental estimation results, while section 7 contains the non-experimental estimation methods and results. Section 7 compares the results from the different methods and provides a discussion. Section 8 concludes.

2 Literature

The effectiveness of active labor market programs has been the topic of a large literature. Reviews are provided by Kluve (2010) and Card et al. (2010), who show that there is some consensus on the effects of different type of programs on post-unemployment outcomes such as employment, wages and job stability. For example, job search assistance is found to be more likely to have positive effects on job finding than alternatives such as public sector employment programs. Several studies have used an experimental approach. Dolton and O’Neill (2002) use random delays in program participation to assess the effect of a job search assistance program in the UK. Van den Berg and Van der Klaauw (2006) analyze a randomized experiment of counseling and monitoring, to show that the program merely shifts job search effort from the informal to the formal search channel. Graversen and van Ours (2008) evaluate an intensive activation program in Denmark, using a randomized experiment. Card et al. (2011) show estimates of the effect of a training program offered to a random sample of applicants in the Dominican Republic. Behagel et al. (2014) perform a large controlled experiment, randomizing job seekers across publicly and privately provided counseling programs. Alternatively, a quasi-experimental approach can be used. Van den Berg et al. (2010) discuss the use of regression discontinuity with duration data and apply their approach to the introduction of the New Deal for the Young People in the UK. Van der Klaauw and Van Ours (2013) analyse the effect of both an employment bonus and sanctions, exploiting policies changes in the bonus levels. Cockx and Dejemeppe (2012) use an regression discontinuity approach to estimate the effect of extra monitoring in Belgium. Other studies use non-experimental
methods such as matching. For example, Brodaty et al. (2002) apply a matching estimator to estimate the effect of activation programs for long-term unemployed workers in France, Sianesi (2004) investigates different effects of active labor market programs in Sweden and Lechner et al. (2011) looks at long-run effects of training programs in Germany. Recently, some studies have used the timing-of-events model (Abbring and Van den Berg (2003)). For example, this model is used to evaluate the effect of benefit sanctions in the Netherlands by Abbring et al. (2005) and to evaluate a French training program by Crépon et al. (2012).

Some studies have compared outcomes of different methods. The before mentioned paper by LaLonde (1986) shows that non-experimental estimators produce results that do not concur with those from experimental evidence for an job training program. Dehejia and Wahba (1999) compare the same experimental findings with results from several matching estimators and shows that these are quite similar. It is shown by Smith and Todd (2005) that these matching results are not robust to different specifications and to the use of different samples and different sets of covariates. More recently, Lalive et al. (2008) evaluate the effect of activation programs offered in Switzerland and compare matching estimators with estimates from a timing-of-events model. Mueser et al. (2007) use a wide set of matching estimators to estimate the earnings effect of job training programs in the US, and compare their estimates to those based on a randomized experiment (Orr et al. (1996)). Kastoryano and Van der Klaauw (2011) evaluate job search assistance for unemployed primary school teachers, where the program assignment rule is known. They compare results from timing-of-events, matching and regression discontinuity methods. Biewen et al. (2012) provide different estimates of the effect of publicly sponsored training programs, while also investigating the sensitivity of their results to certain data features and methodological choices.

3 Institutional setting and the policy discontinuity

In this section we briefly describe the institutional setting and the policy discontinuity in program provision.

In the Netherlands unemployment insurance (UI) is organized at the nationwide level. The UI administration (UWV) pays benefits to all eligible workers. Since October 2006 the eligibility criteria are as follows (see van der Klaauw and de Groot (2014) for an extensive discussion). A worker should have involuntary lost at least half of her working hours with a minimum of five hours per week. The worker should
have worked for at least 26 weeks out of the last 36 weeks. This weeks condition provides a eligibility of three months. During the first two months benefits are 75% of the previous wage, capped at a daily maximum. From the third month onwards it is 70%. If the worker has worked at least 4 out of the last 5 years, the benefits eligibility period is extended with one month for each additional year of employment. The maximum UI benefits duration is 38 months.

A UI benefits recipient is required to register at the public employment office, and to search for work actively. The latter requires making at least one job application each week. Case workers at the UWV provide basic job search assistance through individual meetings, as well as some additional assistance programs. Benefit recipients are obliged to accept any 'suitable job offer'. During the first 6 months a suitable job is defined as a job at the same level as the previous job, between 6 and 12 months it can be a job below this level, and after 12 months any job is suitable and should be accepted. Case workers are responsible for the monitoring of these obligations. In general the intensity of meetings is low though (only in case the case worker suspects that a recipient is unable to find work without assistance a meeting is scheduled). In 2009, each case worker had the possibility of assigning an individual to a range of programs aiming at increasing the job finding rate, if she judged that the job seeker required more than the usual guidance. A large diversity of programs existed, including job search assistance, vacancy referral, training in writing application letters and CV’s, wage subsidies, subsidized employment in the public sector and schooling. Some of these were provided internally by the UWV itself, while others were purchased externally from private companies. Our analysis focusses on the externally provided programs. These can be broadly classified as (with relative frequency in parentheses) job search assistance programs (56%), training or schooling (31%), subsidized employment (2%) or other (11%). Though guidelines existed for the eligibility of job seekers for certain programs, case workers had a large degree of discretion with respect to the decision of which program should be assigned.

This caused the ordering of programs to be insufficiently selective in 2009 and the first three months of 2010, such that the entire budget for 2010 had been depleted by March. In a letter to parliament on March 15 (J.P.H. Donner, Minister of Social Affairs (2010)), the minister of social affairs declared that no new programs should be purchased from that moment onwards. An exception was made for a small number of specially targeted programs (mostly for long-term unemployed) and assistance offered internally by UWV. In section 4 we show that indeed the number of new programs dropped to almost zero in March 2010 and remains very low for the rest
of 2010. This creates a strong discontinuity in the probability of being assigned to a program, which can be exploited to estimate the effect on job finding.

4 Data

We use a large administrative dataset provided by the UI administration, containing all individuals who started collecting UI benefits between April 2008 and September 2010 in the Netherlands. The dataset contains 671,743 observations. We exclude 35,671 spells from individuals employed in a government sector and 17,577 from individuals aged 60 or older. Next we drop 533 spells from individuals working more than 60 hours or less than 12 hours in their previous job and 151 spells from individuals who were eligible for a so-called ‘education and development fund’. Finally we exclude 8518 spells with a duration of zero days and 290 spells from individuals with inconsistent or missing data (such as a negative unemployment duration). The remaining sample contains 608,998 observations (each UI spell is considered an observation, though for some individuals there are multiple observations).

From these data we select a sample that is used for our analysis. There are two reasons for selecting a subsample as opposed to using all observations. First, in the light of identifying the effect of program participation, we prefer a subsample in which pre-reform program participation was most likely. Second, estimates will be more precise when using a more homogeneous sample. Taking these two considerations into account we select all native males, aged 40 to 60 years, with a low unemployment history and belonging to the upper 60% of the income distribution in our data. This leaves us with 116,866 observations. An additional advantage of selecting this sample, is that using a more homogeneous sample improves the performance of the matching estimators that we apply in section 7.1 (Imbens (2014)).

For each spell we observe the day of starting receiving UI benefits and, if the spell is not right-censored, the last day and the reason for the end of the benefit payments. Right censoring occurs on January 1st, 2012, implying after at least 16 months of benefits duration. Furthermore the dataset contains a detailed description of all activation programs (both internally and externally provided) that the individual participated in. Also individual characteristics and pre-spell labor market characteristics are included in the dataset.

Figure 1 shows how the monthly number of individuals entering UI evolves over time. There is a substantial increase in the inflow of unemployed starting December 2008 as a result of the economic crisis. The inflow increases from 2000 to 5000 per month and remained high until the end of 2009. From 2010 onwards the inflow
Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>April 2008</th>
<th>April 2009</th>
<th>April 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment duration (median, days)</td>
<td>245</td>
<td>175</td>
<td>280</td>
<td>275</td>
</tr>
<tr>
<td>Reason for exit (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td>57.2</td>
<td>52.3</td>
<td>52.5</td>
<td>62.2</td>
</tr>
<tr>
<td>End of entitlement period</td>
<td>15.4</td>
<td>21.5</td>
<td>19.1</td>
<td>8.3</td>
</tr>
<tr>
<td>Sickness/Disability</td>
<td>6.7</td>
<td>6.2</td>
<td>7.3</td>
<td>7.6</td>
</tr>
<tr>
<td>Unknown</td>
<td>12.9</td>
<td>12.2</td>
<td>12.8</td>
<td>12.6</td>
</tr>
<tr>
<td>Other</td>
<td>7.8</td>
<td>7.8</td>
<td>8.3</td>
<td>9.3</td>
</tr>
</tbody>
</table>

| Participation external program (%) |             |            |            |            |
| Any program                      | 18.7        | 24.0       | 32.3       | 0.7        |
| Job search assistance            | 11.0        | 17.1       | 21.3       | 0.3        |
| Training                        | 6.0         | 7.3        | 9.7        | 0.2        |
| Subsidized employment            | 0.4         | 0.3        | 0.8        | 0.0        |
| Other                           | 4.9         | 4.4        | 7.7        | 0.2        |

| Participation internal program (%) |             |            |            |            |
| Any program                      | 36.8        | 13.5*      | 40.8       | 39.2       |
| Job search assistance            | 11.6        | 1.7*       | 12.1       | 13.6       |
| Subsidized employment            | 3.2         | 1.1*       | 3.9        | 3.8        |
| Tests                           | 9.7         | 1.9*       | 10.3       | 10.9       |
| Workshops entrepreneurship       | 4.4         | 2.3*       | 6.3        | 3.4        |
| Other                           | 19.7        | 9.1*       | 23.4       | 19.6       |

| Gender (% males) ↑              |            |            |            |            |
| Immigrant (%)↑                  | 0.0         | 0.0        | 0.0        | 0.0        |

| Previous income (%)             |             |            |            |            |
| Low                             | 0.0         | 0.0        | 0.0        | 0.0        |
| Middle                          | 57.4        | 53.8       | 58.3       | 53.1       |
| High                            | 42.6        | 46.2       | 41.7       | 46.9       |
| Age                             | 48.7        | 49.0       | 48.6       | 48.9       |
| Unemployment size (hours)       | 37.2        | 37.1       | 37.4       | 37.3       |
| UI history last 3 year (%)      | 29.2        | 33.8       | 28.8       | 23.3       |
| Education (%)                   |             |            |            |            |
| Low                             | 22.8        | 20.0       | 21.6       | 20.5       |
| Middle                          | 46.5        | 43.0       | 45.7       | 47.1       |
| High                            | 30.7        | 37.1       | 32.7       | 32.4       |
| Observations                    | 116,866     | 1774       | 4441       | 4505       |

Job search assistance contains 'IRO' (Individual reintegration agreement), 'Jobhunting' and 'Application letter'. Training contains 'Short Training' and 'Schooling'. Subsidized employment contains 'Learn-work positions'.

* These values are biased downwards, because participation in internal programs was hardly ever recorded before Jan 2009.

† These variables were used to select the sample.
Figure 1: Entrants in UI benefits per month

Figure 2: Number of UI benefits recipients entering the external programs per month
decreases again. Table 1 presents summary statistics for the full sample, as well as four groups defined by their month of inflow into unemployment. Column (1) shows that for the full sample the median duration of unemployment is 245 days (around 8 months). Almost 60% of those exiting UI find work, while 15% reaches the end of their benefits entitlement period. Almost 7% leaves unemployment due to sickness or disability, the rest leaves due to other reasons or the reason for exit is unknown. In the analysis we focus on exits to employment and exits to unknown.

In the full sample about 19% participates in one of the externally provided programs. Two-thirds of these programs focus on job search assistance, a third involves some sort of training, while only a very small fraction is subsidized employment. About 37% of all individuals participate in an internal program, of which the majority is either some test (such as a competencies test) or job search assistance.

The dataset contains a large set of individual characteristics, including gender, age, immigrant status, education level, previous income, unemployment size, occupation in previous job, unemployment history, geographical region and industry. In the lower panel of Table 1 mean values are presented for some characteristics. The average individual is almost fulltime unemployed (37.2 hours) and 29% has been unemployed in the previous three years. The mean of some of these variables is shown in the lower panel of Table 1.

In columns (2), (3) and (4) the same statistics are presented for three groups of individuals entering unemployment in April 2008, April 2009 and April 2010 respectively. The impact of the policy discontinuity in March 2010 becomes clear from the share of the April 2010 group that participates in an external program. It drops to almost zero for all external programs. To illustrate the impact of the discontinuity in March 2010, we show the number of external programs started per month in Figure 2. The dashed line indicates the policy change in March 2010. The number of programs drops to almost zero in April 2010. Separate graphs for each program are included in the appendix (Figure 17) and show that the discontinuity occurs for each type of program.

The calendar date of entry in UI determines how the policy change affected individuals. Figure 3 shows for the different inflow cohorts the weekly probability of starting an external program. Each cohort reaches the policy discontinuity at a different moment in their UI spell. This is illustrated by the fact that each subsequent cohort experiences the drop in the program entry hazard one month earlier in their unemployment duration. The cohort of March 2010 has a probability of entering

\footnote{The graph shows a smoothed version of the estimated hazard rate into the first external program of each individual.}
an external program close to zero. Figure 3 also shows that participation in some program is, in general, not restricted to a certain duration, though the hazard is rising over the unemployment spell. Before the policy discontinuity the hazards of the different cohorts are very similar, indicating that there are no other major policy changes.

A concern might be that UWV employees have responded to the inability to assign unemployed to external programs, by scaling up the service provided internally. As mentioned, several short programs are offered by UWV employees. There was no increase in available resources for internal assistance, limiting the scope for such an increase. Still, we check for any sudden changes around March 2010. The number of started internal programs per month is shown in Figure 4. Internal programs have only been recorded from 2009 onwards, however this should still provide a sufficiently long series to signal any changes in the pattern around March 2010 (indicated by the striped line). There is no indication of a response around the date of the discontinuity. Separate graphs by type of program are provided in the appendix in Figure 19. The hazard rate into an internal program for different cohorts is shown in Figure 5. The hazards for the cohorts are very similar, supporting the assumption that internal program provision was unaffected by the policy change.
Figure 4: Distribution of starting dates of internally provided programs

![Figure 4: Distribution of starting dates of internally provided programs](image)

Figure 5: Hazard rate into internal programs by month of inflow

![Figure 5: Hazard rate into internal programs by month of inflow](image)
A further concern might be that even though the number of internal programs was not changed in response to the termination of external programs, case workers may have reacted to the unavailability of external programs by shifting their internal programs to these individuals that might otherwise have participated in external programs. This would imply that the policy does not change external program participation to no participation, but, for some individuals, changes it to internal program participation. We argue that if this is the case, this is part of the policy relevant effect. Any unemployment insurance administration would be expected to relocate its remaining resources optimally when a policy instrument becomes unavailable. As such we are interested in the effect of the full policy of external program provision, rather than the effect of one particular program.

However, from an external validity point of view, it is interesting to investigate whether this shift of internal program targeting was indeed happening, in order to more clearly define the treatment we are evaluating. Figure 6 presents mean values of characteristics for individuals starting an internal program per month. If the composition of the group of internal program participants changes after March 2010, this would be an indication that case workers shift their resources towards different types of people. Mean age and weekly hours of unemployment are shown in panel (a) of Figure 6, unemployment and disability history and education level are shown in panel (b), previous income is shown in panel (c) and the share of six industry categories is shown in panel (d). None of the graphs indicate any kind of discontinuity around March 2010.

5 Treatment effects

In this section we define the relevant treatment effects and discuss how they are identified from the policy change.

5.1 Defining the treatment effect

Recall that only a small share of all unemployed workers enters an external program during their unemployment spell. Due to the selectivity of treatment participation, the composition of the program participants differs from the non-participants. The treatment effect for this specific subsample is of most interest. Therefore we focus on the average treatment effect on the treated (ATET), which is defined in the literature as the mean difference between potential outcomes under treatment ($Y_1$) and non-treatment ($Y_0$) for those that receive treatment (see for example Heckman...
et al. (1999)):

\[ ATET = E(Y_1 - Y_0 | D = 1) \]  

\[ D \] is a dummy denoting treatment status. Our key outcome of interest is duration until employment, denoted by \( T \), which is a random variable with cumulative distribution \( F(t) \) and density \( f(t) \). Define \( Y_t = \mathbb{1}(T > t) \), a variable equal to 1 if the individual is still unemployed in period \( t \), and 0 otherwise. The distribution depends on participation in some active labor market program. Denote the duration at which treatment occurs by \( s \), with unemployment duration as a function of treatment, \( T(s) \). The dynamic nature of duration data implies that even for a binary treatment many different treatment effects arise. The treatment can start at different durations, while the effect can be measured at different points in time after the treatment (see for an extensive discussion of dynamic treatment effects Abbring and Heckman (2007)). Using the potential outcomes framework, we have:
\[ Y_{1,t}^*(s) = \begin{cases} 1 & \text{if } T(s) > t \\ 0 & \text{if } T(s) < t \end{cases} \]

The potential outcome under no treatment is

\[ Y_{0,t}^* = \lim_{s \to \infty} Y_{1,t}^* \]

Individuals leave the sample at different durations, such that the composition of the survivors changes over time. This issue, known as dynamic selection, necessitates defining the subgroup for which the treatment effect is defined. As shown by Van den Berg et al. (2010), several possibilities arise because the selectivity also depends on the treatment status. We are interested in the difference in the survival probabilities for treatment \( S = s \) compared to no treatment \( (S = \infty) \), and thus condition on treated survivors. This effect is defined by Van den Berg et al. (2010) as the average treatment effect on the survivors at time \( t \) \( (ATS(s,t)) \), with as arguments the timing of treatment and the timing of the outcome.

\[ ATS(s,t) = E \left[ Y_{1,t}^*(s) - Y_{0,t}^* \mid T > s \right] \]

Analogous to the ATET in the static treatment evaluation, the treatment effect in the dynamic model differs depending on the subgroup for which the effect is defined. The effect of interest is on the subgroup that actually received treatment such that, similar to equation (1), the effect is conditioned on \( S = s \). Thus, the effect we are interested is the average treatment effect on the treated survivors:

\[ ATTS(s,t|S=s) = E \left[ Y_{1,t}^*(s) - Y_{0,t}^* \mid T > s, S = s \right] \] (2)

With the policy discontinuity in mind, we define another treatment effect here, which is the intention-to-treat effect. It is the difference in outflow probability between two subgroups with different treatment rates after \( t = t_2 \).

\[ ITS(t_2,t) = E_{cohort1} \left[ Y_t^* \mid T > t_2, S > t_2 \right] - E_{cohort2} \left[ Y_t^* \mid T > t_2, S \geq t_2 \right] \] (3)

The expectations are taken over the particular subgroup (which is defined as a cohort here). The particular values that we pick for the timing of treatment \( (s) \), the timing at which we measure the outcome \( (t) \) and the particular cohorts will be discussed.
in section 6. First we describe how we can exploit the policy change to identify the ATTS.

5.2 Identification

Estimating the ATTS (equation (2)) without making parametric assumptions is generally not possible from observational data (Abbring and Van den Berg (2005)). A discontinuity in the treatment policy provides an exception and allows estimation of the ATTS. Consider two cohorts. The first enters unemployment at some point in time, with the time until the policy changes equal to $t_1$. The second cohort enters unemployment at some later time, but also before the policy change, and has time until the policy changes equal to $t_2 < t_1$ (see Figure 7). The two cohorts face the same policy of potential program assignment up to a duration of $t_2$, meaning that dynamic selection is the same up to this point. After $t_2$, the first cohort faces another period of potential program assignment, with length $t_1 - t_2$, while cohort 2 faces no potential treatment. As a result, we can compare the outflow to employment in the two cohorts, for those individuals that survived up to $t_2$ and did not start a program up to $t_2$. Any differences in employment for these groups can be assigned to the treatment. Define the survival probabilities as $\bar{F}(t) = 1 - F(t)$. Since only a share of all survivors in cohort 1 receives treatment after $t_2$, comparing the cohorts
identifies the intention to treat effect on the survivors (ITS, equation 3):\(^2\)

\[
\hat{I}TS(t_2, t) = \hat{F}_{\text{cohort1}}(t|T > t_2, S \geq t_2) - \hat{F}_{\text{cohort2}}(t|T > t_2, S \geq t_2)
\]  

(4)

The average treatment effect on the treated follows from dividing the intention to treat effect by the the share of the survivors at \(t_2\) that receive treatment \(S \in [t_2, t_1)\):

\[
\hat{ATTS}(t_2, t) = \frac{\hat{I}TS(t_2, t)}{\sum I(T \geq t_2, S \in [t_2, t_1))} 
\]  

(5)

This estimator provides an average over the treatment effect of treatment at durations \(t \in (t_1, t_2)\). Note that the ATTS as defined in equation (5) measures the ex-post effect of program participation. It imposes that treatment only affects outcomes after it starts. If individuals could anticipate program participation before \(s\), this could violate the no-anticipation assumption (Abbring and Van den Berg (2003)). This is unlikely though for the programs we discuss here. Programs are assigned by caseworkers on an individual basis with mutual consent. There are no strict criteria for participation, so it is impossible for job seekers to know in advance when they will start a program.

Also anticipation of the policy discontinuity that we exploit should be considered. For example, if individuals anticipate the policy change, search behavior (or case worker behavior) may be different before March 2010, thereby biasing the effect estimate. The policy change we are investigating however has the advantage that by definition it was unexpected (because otherwise it would have been prevented in the first place). Since not even unemployment administration was expecting the change, we can safely assume that job seekers and case workers did not anticipate it either.

5.3 Business cycle, seasonalities and cohort composition

A potential problem that remains is that even if two cohorts are compared that enter unemployment relatively shortly after each other, changes in the labor market

\(^2\)Analogous to the effect on the survival probabilities, the effect on the hazard rate can be defined: \(ATTS_\theta(t_2, t) = \theta_{\text{cohort1}}(t|T \geq t_2) - \theta_{\text{cohort2}}(t|T \geq t_2)\) with \(\theta(t) = \frac{f(t)}{1 - F(t)}\) the hazard rate. The policy change only allows for estimation of \(ATTS_\theta(t_2, t_2)\) though, which is the instantaneous treatment effect on the hazard of the treated survivors. Even in the case of randomized treatment assignment the control and treatment groups are only comparable at the moment of randomization and not at any duration after that, due to different dynamic selection. The same holds for our estimator: the hazards of two cohorts are comparable until \(t_2\), but not after that. An instantaneous effect of the programs is unlikely, since the programs run for some time. Therefore, we decide to use the survival probabilities as the outcome measure.
conditions may lead to differences in outcomes. We discuss how this may affect our estimates, and how we correct for this. Figure 8 presents two aggregate labor market statistics, the unemployment rate and GDP. Both the raw series and the seasonally corrected series are shown. In the period before the policy discontinuity, 2009 and the beginning of 2010, unemployment is rising due to the economic crisis. During 2010 it decreases slightly, while in 2011 it increases again. In the short-run, seasonalities are the main cause of variation in both unemployment and GDP. Fluctuations in labor market conditions may affect outcomes in two ways. First, they affect the inflow of unemployed. For example, as the crisis worsens, different types of workers in terms of education or experience may be laid off. As a result the composition of cohorts varies, which also affects the outflow probabilities. Second, labor market conditions affect outflow probabilities directly, as it is more difficult to find employment when unemployment is high.

To correct estimates for the differences in composition we can exploit the set of covariates available in the data. We compute population weights based on the March 2010 cohort, with respect to three previous income categories, an indicator for having been unemployed in the past three years, an indicator for being married or cohabitating, age categories, an indicator for being part-time unemployed (less than 34 hours per week) and three education categories. Interacting these covariates we have 288 groups. Define the share of group $g$ in cohort $c$ by $\alpha_{c,g}$. The weight
assigned to an observation belonging to group $g$ and cohort $c$ is defined by:

$$w_{c,g} = \frac{\alpha_{mar2010,g}}{\alpha_{c,g}}$$

Each survivor function is then estimated using these weights, thereby correcting for differences in composition. Results do not change a lot by using these weights. In the appendix we also present unweighted results.

The direct effect of business cycle and seasonal effects on employment probabilities requires some more discussion. To formalize these factors, consider the following simple model. Assume that the hazard rate to employment ($h$) for cohort $\tau$ depends on the duration of unemployment ($t$), the effect of the business cycle ($b_\tau$), the effect of seasonalities ($c_\tau$), individual characteristics ($x$), and program participation ($\gamma_\tau(t)$):

$$h(t, \tau, x) = f(t, \tau, x, \gamma)$$

(6)

In order to be able to say something about how to correct for business cycle effects when identifying the effect of program participation, we need to make some assumptions about the hazard. First, by reweighting our estimates we can eliminate individual characteristics $x$. Then, we assume that the business cycle, seasonalities and treatment have an additive effect on the baseline hazard, where each of these impacts may vary by duration $t$. Note that this is very flexible as we do not assume anything on how these factors vary by duration. The duration dependence of the hazard is denoted by $\lambda(t)$ (which is common for all cohorts).

$$h(t, \tau) = \lambda(t) + b_\tau(t) + c_\tau(t) + \gamma(t)$$

(7)

From the hazard rate we can construct the survival function.

$$\bar{F}(t)_\tau = \exp\left(-\int_0^t h(u|\tau)du\right)$$

$$= \exp\left(-\int_0^t \lambda(u)du - \int_0^t b_\tau(u)du - \int_0^t c_\tau(u)du - \int_0^t \gamma(u)du\right)$$

(8)

Taking the logarithm of the survival function we have:

$$\log \bar{F}_\tau(t) = -\int_0^t \lambda(u)du - \int_0^t b_\tau(u)du - \int_0^t c_\tau(u)du - \int_0^t \gamma(u)du$$

$$\equiv \Delta(t) + B_\tau(t) + C_\tau(t) + \Gamma(t)$$

(9)
This implies that the business cycle, seasonalities and the treatment have additive impacts on the log of the survival function. For ease of exposition, assume there is a binary treatment at the cohort level that can start at any duration. Seasonalities are by definition those factors that are common across different years, such that \(C_\tau(t) = C_{\tau-12}(t)\), for all \(\tau\). The intention-to-treat effect estimator defined by (3) compares the survivor functions of two cohorts. For example, consider a comparison of the January 2010 cohort with the October 2009 cohort. If instead of survivor functions we compare the log of survivor functions, this identifies the treatment effect plus the difference in seasonal and business cycle effects between the two cohorts:

\[
\Delta^1(t) = \log \bar{F}_{j10}(t) - \log \bar{F}_{o09}(t) = \Gamma(t) + [C_{j10}(t) - C_{o09}(t)] + [B_{j10}(t) - B_{o09}(t)]
\] (10)

How large these biases are depends on the length of the time interval between the two cohorts, and the stability of the labor market. The effect of the additional treatment starts at duration \(t = t_2\) (as explained in Figure 7). This implies that differences between the cohorts before \(t_2\) are not caused by the treatment (as discussed in section 5.2). Such differences are either due to the seasonal and business cycle terms, or they are simply noise in the data. Since the survival function integrates over the hazard rate, such differences at early duration would end up in the treatment effect estimated at later durations. To prevent this, we can estimate the survival function conditional on survival up to duration \(t_2\). Conditioning on \(T > t_2\) implies that all integrals in (8) run from \(t_2\) to \(t\) instead of 0 to \(t\).

We can potentially improve on this estimator by applying a difference-in-differences approach. By subtracting the same cohort difference from a year earlier, we eliminate the seasonal effects, at the cost of adding extra business cycle effects:

\[
\Delta^2(t) = [\log \bar{F}_{j10}(t) - \log \bar{F}_{o09}(t)] - [\log \bar{F}_{j09}(t) - \log \bar{F}_{o08}(t)] = \Gamma(t) + [B_{j10}(t) - B_{o09}(t)] + [B_{j09}(t) - B_{o08}(t)]
\] (11)

Whether this is preferable over \(\Delta^1(t)\) depends on the relative sizes of the business cycle and seasonal effects. If the interval between the cohorts is sufficiently small (as in this particular example), Figure 8 suggests that seasonal effects are larger,

\[3\]We are considering the effect of a treatment on the cohort survivor function here. Therefore the treatment can be interpreted as program participation of a share of the cohort at a particular duration, which is exactly the case in our setting.

\[4\]And also conditional on no treatment before \(t_2\).
such that $\Delta^2(t)$ is preferable. If one believes that business cycle differences are small enough to ignore when the interval is small, this is a satisfactory estimator of the effect of treatment $\Gamma(t)$.

In the next section we provide estimates of the two estimators.

6 Experimental analysis

Applying the estimators defined in section 5.2 requires choosing which cohorts to compare. We define a cohort as all individuals entering unemployment within one month. The time between cohorts should be rather small, to minimize business cycle and seasonal effects. There is a tradeoff though, since the more time there is between cohorts, the larger is the difference in time of exposure to potential treatment. We decide to use cohorts three months apart. Second, the cohorts should not enter unemployment too long before the discontinuity, since for those cohorts a large share will already have exited when the policy changes. Therefore we choose to use each of the cohorts of October 2009 until January 2010, facing between five and two months of potential treatment respectively. Each of these cohorts will be compared to the cohort entering unemployment three months earlier. The survivor function of each of these cohorts is presented in Figure 9. Around 50% of the UI benefits
recipients finds employment within 12 months and there are quite some differences between the four cohorts.

6.1 Intention-to-treat effect

We take the difference between each survivor function and the survivor function of the cohort entering unemployment three months earlier ($\Delta^1(t)$). As discussed in section 5.2, we condition on survival and no-treatment up to the duration at which the later cohort reaches the policy discontinuity. So when comparing January 2010 with October 2009, only individuals are included with an unemployment duration of at least three months, who do not start an external program in the first two months. The differences up to a duration of 18 months are presented in panel (a) of Figure 10. We find a negative effect on job finding over the first few months of around 4%-points in all but one comparison. After about 10-12 months the difference disappears and all estimates are close to zero.

These estimates are based on simple differences between cohorts, thus not taking into account fluctuations in labor market conditions. By subtracting the same differences from a year earlier, we correct for cohort differences that are constant across years (such as seasonalities). Estimates from such a difference-in-differences approach ($\Delta^2(t)$) are presented in panel (b) of Figure 10.\(^5\) Again we find a negative effect on job finding in the first months. At longer durations, the estimates diverge somewhat. In the appendix we present each line with a 95% confidence interval (standard errors are computed with bootstrapping) in Figure 20. The early negative effect is always significantly different from zero, while none of the estimates at longer durations are significantly different from zero. Note that each comparison measures the effect of additional treatment at a slightly different duration. For example, the January 2010-October 2009 comparison measures the effect of additional treatment in the 4th-6th month of unemployment, while the December 2009-September 2009 comparison measures the effect of additional treatment in the 5th-7th month of unemployment.

The results show a pattern that is quite consistent across different cohort comparisons and across the two estimators. Job finding is significantly reduced in the early months, while the difference disappears after 6-12 months. This finding is line with the lockin effect often found in the literature. When a program starts, participants shift attention from job search to the program and actually have a lower

\(^5\)When estimating $\Delta^2(t)$ we only present estimates up to the duration at which the cohorts from a year earlier reach the discontinuity, which is between 15 and 18 months. Estimates at longer durations would be biased as the earlier cohorts would be affected by the discontinuity.
Figure 10: Treatment effect estimates, conditioned on $T > t_2, S > t_2$

![Graphs showing treatment effect estimates](image)

(a) $\Delta^1(t)$
(b) $\Delta^2(t)$

job finding rate. This negative effect disappears after some months, but we do not find any (positive) effects at longer durations. All these findings are intention-to-treat effects. To estimate the average effect of treatment on individual employment probabilities they need to be scaled by the differences in treatment rates.

### 6.2 Average treatment effect

To interpret the estimates as average treatment effect, we divide each estimate by the difference in treatment share between the cohorts that are being compared (as defined in equation (5)). Since we observe the date at which a program starts, we can compute estimates of the difference in treatment share, analogous to the three months differences for job finding. Define $\bar{F}_{treatment}$ as the survivor function for treatment, where an exit occurs at the start of the first program. The differences in treatment rates (for the January 2010 - October 2009 comparison) are given by:

$$\Delta^1_{Treatment}(t) = \bar{F}_{Jan10}^t | T, S > 2 \text{ months} - \bar{F}_{Oct09}^t | T, S > 2 \text{ months}$$ (12)

And similar for the other cohorts that are compared. These estimates are presented in panel (a) of Figure 11. We find a clear increase as soon as the first cohort reaches March 2010. The difference increases for approximately three months, after which the comparison cohort reaches March 2010. From that point onwards, both cohorts receive no treatment, and the difference remains stable. The difference is between 15 and 20 % points. Note that the difference in treatment rate increases as we compare cohorts entering longer before the discontinuity. For earlier cohorts, the three months of extra treatment occur at a later duration. Since the proba-
Figure 11: Differences in treatment rates, conditioned on $T > t_2$ and $S > t_2$

The probability of a participating in a program increases with duration (see Figure 3), the number of extra programs is higher for these cohorts. The difference in treatment share can also be computed using the same difference-in-differences approach as in $\Delta^2(t)$. Such estimates are presented in panel (b) of Figure 11. Due to differencing with cohorts from a year earlier, the differences are somewhat less smooth, though the main pattern remains. Confidence intervals are presented in Figure 21 in the Appendix. All differences are highly significant.

The results of dividing the estimates from panel (a) in Figure 10 by the treatment difference are presented in panel (a) of Figure 12. The pattern does not differ much from that of the intention-to-treat effects, though quantitatively the effects are much larger. Program participation reduces the job finding probability during the first 2-3 months by about 40 %-point, while after 10 months employment probabilities are similar again and there is no significant effect. Difference-in-differences estimates ($\Delta^2(t)$) are presented in panel (b) of Figure 15. The pattern is quite similar. There is a negative effect on the job finding probability of 40%-point directly after program participation starts, which decreases in magnitude over time towards a zero effect after about 8 months (confidence interval are presented in Figure 22 in the Appendix).

6.3 Common trend assumption

The assumption that the business cycle terms in (11) are negligible is equivalent to the common trend assumption in a standard difference-in-differences estimator. It requires that in the absence of the policy discontinuity, the difference in employment rate between the January and November cohort would have been the same in 2009/2010 as in 2009/2008. This is by definition untestable. However, we can
get an indication of the plausibility of the assumption, by investigating the survivor functions over the first months of each cohort. All estimators condition on survival up to \( t_2 \), but we can use information on job finding before \( t_2 \) to test whether the common trend assumption holds in these first months. It requires a sufficient number of pre-discontinuity months in the latest cohort, such that we can only do so for the comparisons of December 2009, November 2009 and October 2009. Basically, we estimate \( \Delta^1(t) \) and \( \Delta^2(t) \) for \( t \leq t_2 \), without conditioning on survival up to a certain duration. If we find significant differences, this suggests that the common trend assumption might be violated.

Estimates are presented in Figure 13 for December 2009 - September 2009, November 2009 - August 2009 and October 2009 - July 2009, including 95% confidence intervals, that were computed using bootstrapping. We find that all estimates are significantly negative, though small in magnitude. This implies that the treatment effect estimates are biased downwards somewhat, and thus provide lower bounds of the effect.
7 Non-experimental methods

Next, we turn to non-experimental methods to estimate the effect of the activation programs on job finding. We apply a matching estimator in subsection 7.1 and estimate the timing-of-events model in subsection 7.2.

7.1 Matching

Matching methods include a set of popular estimators to compute counterfactual outcomes in the absence of randomized assignment to treatment (Imbens and Wooldridge (2009), Abadie and Imbens (2011)) and date back to Rosenbaum and Rubin (1983). The approach relies on two main assumptions. First, selection into treatment is on observables only:

\[ Y_{0,t}^*(s), Y_{1,t}^*(s) \perp S \mid X \]  

(13)

Also referred to as the unconfoundedness assumption, it implies that after conditioning on a set of observed characteristics, assignment to treatment is independent of the potential outcomes. If selection into treatment is on observable characteristics only, a counterfactual outcome can be constructed using different methods that correct for the differences in observed characteristics. Matching is closely related to a simple OLS framework which can also correct for differences in observable characteristics and relies on a similar unconfoundedness condition. However, in an OLS regression this assumption is combined with strong functional form assumptions (covariates enter the model additively and linearly), such that the model is much more restrictive (Imbens (2014)).

If treatment selection occurs also on non-observables, matching estimators yield biased results. The bias of matching estimators depends on the degree of selection that is captured by observed characteristics. Including a rich set of covariates is thus crucial. It is often argued that employment histories are particularly important (see for example Card and Sullivan (1988), Heckman et al. (1999), Gerfin and Lechner (2002), Lechner et al. (2011)). Our administrative data set allows the use of a rich set of covariates. We include individual characteristics (age, education level, marital status, region), employment history variables (previous income, unemployment history, previous industry) and current unemployment spell variables (unemployment size in hours, a sickness or disability dummy, maximum UI eligibility). As such we are confident that we include a set of covariates that includes all information that is generally available in evaluations of active labor market programs. Still, it is difficult
to exclude a priori that selection also depends on non-observable characteristics. By comparing the matching estimates to our baseline (experimental) estimates from section 6, we aim to assess to what extent such a bias plays a role.

The second condition for applying matching is that there is a common support in the distribution of the covariates between treated and non-treated individuals. Let $D_i$ be an indicator equal to one for individuals participating in the program. The common trend assumption states:

$$0 < \Pr(D = 1 | X = x) < 1 \; \forall x$$

It implies that, in order to match treated individuals with untreated individuals who are similar with respect to covariates, a complete overlap in the distribution of these covariates is required between the two groups. A simple way of investigating the plausibility of this assumption is to estimate a propensity score model and plot the density of predicted values for treated and non-treated individuals. The overlap assumption is likely to be violated in case there is little overlap in the two densities, or if there is a lot of mass close to zero or one in one of the groups (Busso et al. (2009)). In 23 in the appendix we present the predicted propensity scores for treated and non-treated individuals, after estimating a logit model for the likelihood of being treated. The plots suggest that there is sufficient overlap: both densities have most mass between 0.05 and 0.3 and there is hardly any mass close to zero or one.

### 7.1.1 Sample and treatment definition

We start by defining the treatment and the relevant sample. Our aim is to assess the performance of matching estimators, by comparing results to those of the experimental estimator. The sample should overlap largely with that used in the previous section, such that the distribution of individual characteristics and labour market conditions are the same. Therefore we include individuals entering unemployment between September and December 2009. Recall that our experimental estimates are based on comparisons of the cohorts July 2009 up to and including January 2010. So there is a large overlap in the samples. In order to estimate the same treatment effect as in our non-experimental analysis, we define treatment as being selected for program participation in months 3-5 and exclude all individual that leave unemployment before month 3 or those that start a program before month 3. This corresponds to the effect that we measure in the January 2010 - October 2009 comparison in Figure 12. All other individuals are classified as belonging to the control group, however we censor their observation if they start a program after month 5.
In the next section we provide the details of this approach, which is similar to that used by Lalive et al. (2008). From the perspective of assessing the performance of the matching approach, the estimates should not be based on a sample in which treatment assignment is affected by the policy discontinuity. The sample that we use might be slightly affected. The impact on the estimates is small though, because we censor all control group observations if they start a program. As a robustness check we show that using a sample consisting of the same monthly cohorts from 2008 instead of 2009 gives very similar estimates.

Figure 14: Average treatment effect on the treated (with 95% confidence intervals): Matching estimator using Kaplan-Meier survival functions

7.1.2 Approach and estimation results

Standard static matching estimators require excluding individuals who start treatment after month 5 from the pool of potential controls. If they would be included, it would change the interpretation of the estimate into the effect of early treatment relative to later treatment, which is not the parameter we are interested in in this study. Excluding these observations is problematic as well, as it creates a selective control group. To address this problem we take the following approach. First, we use a logistic model to predict for each individuals the propensity score, classifying those that start treatment in months 3-5 as the treated group, and all others as the control group.\footnote{All individuals leaving unemployment before month 3 or starting treatment before month 3 are excluded, which was also done in the experimental approach.}

As a second step, we estimate for both the treated and the matched control group the Kaplan-Meier survival functions, using as weights the inverse of the predicted propensity scores.\footnote{For control group individuals the weights are given by $\hat{p}(X_i)$, the predicted probability of being
month 5 are censored at the start of their program. This way, we can include all potential control observations, however we do not use information about job finding after a control group individual starts treatment. The difference between the survival functions provides an estimate of the average treatment effect on the treated. Standard errors are computed using bootstrapping. The estimates, including a 95% confidence interval are presented in panel (a) of Figure 14. The estimates of the effect of program participation on the job finding probability are significantly negative at any duration, though the negative effect becomes smaller over time. At an unemployment duration of 5 months, those that started treatment during month 3-5 have an almost 20% points lower probability of having found employment. This difference decreases to 10% points after 18 months of unemployment.

As an alternative, we use a dataset including individuals entering unemployment one year earlier. Those are unaffected by the policy discontinuity, however they may face different labor market conditions. Estimates are presented in panel (b) of Figure 14. The results are very similar, showing the same negative effect which decreases in magnitude over time. Standard errors are slightly larger, which is due to the fact that more control group individuals start treatment after month 5 and have to be censored at that point. In the appendix we present a range of other standard matching estimators, showing that the results are also very similar (Figure 24).

We conclude that the estimates are quite robust to the choice of the sample. Each shows a negative estimate of the program effect, which is almost 20% points after 5 months of unemployment, and decreases in magnitude to a negative effect of 5-10% points after 18 months of unemployment. In the next section we apply the timing-of-events approach, which requires more functional form assumptions, but allows to explicitly model unobserved heterogeneity.

### 7.2 Timing of events model

A common model used for evaluating dynamic treatment effects is the timing-of-events model (Abbring and Van den Berg (2003)). It specifies both the outcome model and the treatment assignment model. These models are continuous time duration models, in which the unobserved terms in the two models are allowed to be correlated. Identification of the model relies on the randomness of the exact timing of program participation. The main identifying assumption is the no-anticipation assumption, stating that the job finding rate is not affected by program participation before the program starts. It does not rule out that the treatment probability differs between individuals and that individuals are aware of this. Some job seekers

\[ \text{treated}. \text{ For treated individuals the weights are given by } 1 - \hat{p}(X_i). \]
may have a high probability of program assignment and know this. Only the exact
timing of the program start should be unanticipated. This assumption is there-
fore considerably weaker than the conditional independence assumption (as defined
by equation (13)). The timing of events model on the other hand requires extra
functional form assumptions: the effect of covariates and the treatment on the job
finding rate is assumed to be constant and proportional.

This model has for example been used by Abbring et al. (2005), Berg et al.
(2004), Lalive et al. (2005) and Van der Klaauw and Van Ours (2013). Consider
an individual entering unemployment at calendar date $\tau$. Her baseline job finding
rate depends on the number of days of unemployment ($t$) and is denoted by $w_T(t)$.
Calendar time fixed effects are captured by the function $q_T(\tau_0 + t)$. The hazard rate
further depends on her observed characteristics ($x$) and unobserved characteristics
$v_T$. After starting a program, the hazard rate is shifted by the treatment effect
$\delta$. Since the treatment is likely to affect the hazard differently over duration, we
include a flexible specification with a set of effects, $\delta_n I_{t \in (s_n, s_{n+1})}$. For values of $t$
between $s_n$ and $s_{n+1}$, the indicator function $I_{t \in (s_n, s_{n+1})}$ equals one. The effect of
program participation starts directly after the start of the program ($s_1 = s$). We
allow different effects for the first 6 months ($s_2 = 6$) and the period after 6 months.
The job finding hazard rate is given by:

$$h_T(t|x, \tau_0, s, v_T) = w_T(t)q_T(\tau_0 + t) \exp \left[ x\beta_T + \sum_n \delta_n I_{t \in (s_n, s_{n+1})} \right] v_T$$  \hspace{1cm} (14)

Estimation of equation (14) yields a biased estimate of the treatment effects because
the treatment assignment is non-random (suspectedly even conditional on the ob-
served characteristics) and is thus correlated with the unobserved determinants of
unemployment duration $v_T$. To account for this, treatment assignment is modelled
simultaneously. The hazard rate into treatment is given by:

$$h_S(t|x, \tau_0, v_S) = w_S(t)q_S(\tau_0 + t) \exp(x\beta_S)v_S$$  \hspace{1cm} (15)

With all notation similar to equation (14). The unobserved term $v_S$ is allowed to
be correlated with $v_T$, with joint discrete distribution $g(v_T, v_S)$. We take $g(v_T, v_S)$
to be a bivariate discrete distribution with an unrestricted number of mass points
indexed by $k$. The corresponding probabilities are:

$$p_k = P(v_T = v^k_T, v_S = v^k_S) \hspace{1cm} \forall k$$  \hspace{1cm} (16)

Estimation results show that using $k = 2$ is sufficiently flexible. We parameterize the
duration dependence pattern of the hazards \((w_T(t), w_S(s))\) with a piecewise constant function:

\[
w_j(t) = \sum_{m=1}^{M} \pi_m I_{t_{m-1} \leq t < t_m}
\]

(17)

Where \(I_{t_{m-1} \leq t < t_m}\) is an indicator for duration being between \(t_{m-1}\) and \(t_m\) and \(\pi_m\) the function value. This function is very flexible, as by increasing the number of intervals \(m\) it can approximate any pattern arbitrarily well. We use the following interval boundaries (in months): 0-2, 2-4, 4-6, 6-9, 9-12, 12-18, 18-24, 24-48.

Calendar time effects are also modelled by a piecewise constant function, where each interval is taken to be a quarter-of-a-year.\(^8\) Estimation of the parameters is performed by maximizing the log-likelihood. Subscript individuals \(i\), with \(I\) the number of individuals observed. Some individuals leave unemployment for other reasons than finding employment or remain unemployed until the end of our observation period. Define \(c^{T}_i = 1\) if an exit to work is observed, and \(c^{T}_i = 0\) otherwise. Similarly, we define \(c^{S}_i = 1\) if a program start is observed before the end of unemployment duration and \(c^{S}_i = 0\) otherwise. The log likelihood is then given by:

\[
\log \mathcal{L} = \sum_{i} \log \left[ \sum_{k=1}^{K} p_k h_T(t_i|x_i, \tau_{0i}, s_i, v^k_T) c^{T}_i \exp \left( - \int_{0}^{t_i} h_T(u|x_i, \tau_{0i}, s_i, v^k_T) du \right) \right]
\]

\[
\sum_{i} \log \left[ \sum_{k=1}^{K} p_k h_S(s_i|x_i, \tau_{0i}, v^k_S) c^{S}_i \exp \left( - \int_{0}^{s_i} h_S(u|x_i, \tau_{0i}, v^k_S) du \right) \right]
\]

The log likelihood is maximized over all parameters.

### 7.3 Estimation results

The model is estimated using the full dataset of 116,866 observations, as described in section 4. Full estimation results are presented in the appendix in table 3.\(^9\)

The estimates of the treatment effects are presented in table 2. The effect of program participation is estimated to have a large, significantly negative effect on the job finding rate in the first two months \((\delta_{0-2\text{ months}})\), with the coefficient equal

\(^8\)We normalize the function value for the first interval to 1 in each of the four piecewise constant functions.

\(^9\)In column (1) of Table 3 the coefficients of the job finding hazard (equation (14)) are presented. Most covariates have the expected sign and are highly significant. The oldest age group has a lower job finding rate, as do lower educated individuals. Being (partially) disabled, sick or having an unemployment history also reduces the job finding rate. Furthermore we find that the job finding rate decreases steadily over unemployment duration. As expected there are some fluctuations in the job finding rate related to the calendar time. In column (3) the coefficients of the treatment hazard are presented (equation (15)).
Table 2: Treatment effect estimates Timing-of-Events model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{(0-2 \text{ months})}$</td>
<td>-0.42 0.00</td>
</tr>
<tr>
<td>$\delta_{(3-6 \text{ months})}$</td>
<td>-0.12 0.00</td>
</tr>
<tr>
<td>$\delta_{(\geq 6 \text{ months})}$</td>
<td>0.05 0.00</td>
</tr>
</tbody>
</table>

to -0.42. In the third and fourth month the effect is still significantly negative, but smaller in magnitude (-0.12). After six months ($\delta_{(\geq 6 \text{ months})}$) program participation has a small, significantly positive effect on the probability of finding a job. To interpret the magnitude of these effects, we transform them into an effect on the probability of being employed at some duration $t$ since being unemployed. This can be done by computing the survivor function for an average individual (we take the mean value of each individual characteristic in the sample), entering unemployment in December 2009. We simulate the survivor based on the estimated model twice. Once imposing that the individual never participates in a program, the second time imposing program participation after two months of unemployment. Both survivor functions (up to a duration of two years) are presented in panel (a) of Figure 15, including one standard error bounds. Standard errors are computed using the delta method. Participating in a program after two month of unemployment lowers the probability of being employed subsequently, in accordance with the negative effect estimate. The difference between the survivors is presented in panel (b) of Figure 15. The effect is significantly negative directly after the program starts, and increases in magnitude up to a 5% points difference after 6 months. Afterwards the effect decreases in magnitude and as standard errors increase it is no longer significantly different from zero after about 14 months.

Figure 15: Timing of Events model: Simulated treatment effect

(a) Simulated survivor functions (with one standard error bounds)  
(b) Simulated treatment effect (with 95-% confidence interval)
8 Discussion

We have estimated the impact of activation programs on the exit rate to work using three approaches: exploiting the policy change with a non-parametric difference-in-differences estimator, using a set of matching estimators, and using the timing-of-events model. In this section we compare the different results and discuss reasons for the differences in outcomes.

In section 6, we presented the results of our preferred difference-in-differences specification in panel (b) of Figure 12. The program effects are identified from the discontinuous drop in program provision in March 2010. The figure presents estimates from four cohort comparisons. For the aim of simplicity when comparing the different methods we focus on one set of estimates, which are those obtained from comparing the cohorts of January 2010 and October 2009. For these cohorts the difference in treatment accrues in months 3-6 of unemployment duration. So the effect should be interpreted as the effect of participating in a program after 3-6 months of unemployment on the probability of employment.

In section 7.1, we presented results using matching estimators. These program effect estimates are identified from the conditional independence assumption (equation (13)) and some functional form assumptions. Since the results of the different estimators were extremely robust, it does not matter which one we select here. We focus on the propensity score matching estimator (Figure 14, panel (c)).

In section 7.2 we estimated the timing of events model and obtained an estimate of the program participation effect. Identification of the treatment effect relies on the no anticipation assumption and the functional form of both the outcome and the treatment selection equation. It also imposes that the treatment effect is homogeneous across individuals. By simulating survivor functions, an estimate of the effect on the employment probability was obtained.

In Figure 16 each of the three estimates are presented in one graph. Note that each of the estimators measures the effect of starting participation in a program within months 3-5 as compared to never starting a program on the probability of being employed, averaged over the subgroup that received treatment in the pre-discontinuity period:

Since each method measures the same effect, the estimates should be similar (up to ...sampling....) if the assumptions of the methods are fulfilled. On the contrary, we find differences in the results. Difference-in-differences suggests a negative impact on job finding in the first three months after the program starts and a zero effect on job finding afterwards. The matching estimates are always negative, though decreasing in magnitude as duration since treatment increases. Timing-of-events suggests a
negative effect on job finding in the first 9 months after the program starts, which
decreases in magnitude towards a zero effect after a year. So we find that the
timing-of-events method quite closely resembles the experimental estimates, while
the matching estimators underestimate the effect at the medium run. This suggests
that the matching estimators are unable to adequately correct for the selection into
treatment. The timing-of-events model requires more functional form assumptions
but allows for unobserved heterogeneity.

9 Conclusion

Several methods are available when evaluating treatment effects of activation pro-
grams for job seekers. In this paper we compare estimates from three methods, using
estimates from an experimental method as the baseline. The baseline estimates rely
purely on exogenous variation in program participation, caused by budgetary prob-
lems of the unemployment insurance administration. Exploiting this variation with
a difference-in-differences approach, we find that program participation reduces out-
flow to work in the first couple of months with up to 40%-point, and has no effect
on job finding at longer durations (10-18 months).

We find that the non-experimental methods that we apply, matching and timing-
of-events, lead to slightly different results. Both reproduce the negative short-run
effect, however the matching estimators are negative even after 18 months. The
timing-of-events model estimate converges to zero after 16 months. The discrepancy
between the results suggest that the matching approach is unable to adequately
correct for the selection into treatment such that the identifying assumption is likely to be violated.

We conclude that these non-experimental methods should be used with caution. Especially in the case of activation programs where selectivity of participation is a large issue, even a large set of observed characteristics may not suffice to correct for the bias. Some studies have used even richer data sets including ‘soft’ variables. For example Thomsen et al. (2008) include caseworkers’ subjective assessment as a control and Caliendo et al. (2014) include variables such as personality traits, attitudes, expectations, and job search behavior. A useful next step would be to compare non-experimental methods that include such information with experimental evidence.

References


10 Appendix
Figure 18: Timing of externally provided programs

(a) IRO (Individual Reintegration Agreement)  
(b) Short training  
(c) Jobhunting  
(d) Regular programs  
(e) Learn-work jobs  
(f) Schooling

Table 3: Estimates Timing-of-Events model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient(Jobfinding)</th>
<th>p-value</th>
<th>Coefficient(Treatment)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 45-50</td>
<td>0.05</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Age 50-55</td>
<td>0.05</td>
<td>0.00</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Age 55-60</td>
<td>-0.21</td>
<td>0.00</td>
<td>-0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>Middle educated</td>
<td>0.04</td>
<td>0.00</td>
<td>0.21</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Continued on next page...
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient(Jobfinding)</th>
<th>p-value</th>
<th>Coefficient(Treatment)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High educated</td>
<td>0.06</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.82</td>
</tr>
<tr>
<td>Unemployed hours</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Income (cat. 4)</td>
<td>0.06</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>Income (cat. 5)</td>
<td>-0.00</td>
<td>0.69</td>
<td>-0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>Sickness</td>
<td>-0.28</td>
<td>0.00</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>Disability</td>
<td>-0.47</td>
<td>0.00</td>
<td>0.06</td>
<td>0.31</td>
</tr>
<tr>
<td>Sickness and disability</td>
<td>-0.17</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.55</td>
</tr>
<tr>
<td>Region 1</td>
<td>-0.06</td>
<td>0.02</td>
<td>1.47</td>
<td>0.00</td>
</tr>
<tr>
<td>Region 2</td>
<td>-0.09</td>
<td>0.00</td>
<td>1.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Region 3</td>
<td>-0.06</td>
<td>0.02</td>
<td>1.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Region 4</td>
<td>-0.09</td>
<td>0.00</td>
<td>1.14</td>
<td>0.00</td>
</tr>
<tr>
<td>Region 5</td>
<td>-0.08</td>
<td>0.00</td>
<td>0.84</td>
<td>0.00</td>
</tr>
<tr>
<td>Region 6</td>
<td>-0.09</td>
<td>0.00</td>
<td>0.93</td>
<td>0.00</td>
</tr>
<tr>
<td>UI history (cat. 2)</td>
<td>0.15</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.59</td>
</tr>
<tr>
<td>UI history (cat. 3)</td>
<td>-0.01</td>
<td>0.27</td>
<td>-0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>UI history (cat. 4)</td>
<td>-0.23</td>
<td>0.00</td>
<td>-0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>UI history (cat. 5)</td>
<td>-0.23</td>
<td>0.00</td>
<td>-0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>Marital status 2</td>
<td>0.13</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.76</td>
</tr>
<tr>
<td>Marital status 3</td>
<td>0.24</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.82</td>
</tr>
<tr>
<td>Marital status 4</td>
<td>0.08</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.80</td>
</tr>
<tr>
<td>Marital status 5</td>
<td>0.10</td>
<td>0.00</td>
<td>-0.03</td>
<td>0.47</td>
</tr>
<tr>
<td>Marital status 6</td>
<td>0.04</td>
<td>0.15</td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>UI eligibility</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\delta_{0-2}$</td>
<td>-0.42</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{3-6}$</td>
<td>-0.12</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{\geq 6}$</td>
<td>0.05</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dur. dep. 2</td>
<td>3.06</td>
<td>0.00</td>
<td>1.28</td>
<td>0.00</td>
</tr>
<tr>
<td>Dur. dep. 3</td>
<td>2.31</td>
<td>0.00</td>
<td>1.56</td>
<td>0.00</td>
</tr>
<tr>
<td>Dur. dep. 4</td>
<td>2.00</td>
<td>0.00</td>
<td>2.25</td>
<td>0.00</td>
</tr>
<tr>
<td>Dur. dep. 5</td>
<td>1.50</td>
<td>0.00</td>
<td>2.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Dur. dep. 6</td>
<td>2.49</td>
<td>0.00</td>
<td>1.55</td>
<td>0.00</td>
</tr>
<tr>
<td>Dur. dep. 7</td>
<td>1.89</td>
<td>0.10</td>
<td>3.30</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 2</td>
<td>1.27</td>
<td>0.00</td>
<td>1.23</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 3</td>
<td>0.82</td>
<td>0.00</td>
<td>1.62</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 4</td>
<td>0.91</td>
<td>0.00</td>
<td>1.50</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 5</td>
<td>0.80</td>
<td>0.00</td>
<td>1.75</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 6</td>
<td>0.80</td>
<td>0.00</td>
<td>1.74</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 7</td>
<td>0.58</td>
<td>0.00</td>
<td>2.59</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 8</td>
<td>0.94</td>
<td>0.00</td>
<td>2.05</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 9</td>
<td>0.86</td>
<td>0.00</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 10</td>
<td>1.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 11</td>
<td>0.77</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Continued on next page...*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient(Jobfinding)</th>
<th>p-value</th>
<th>Coefficient(Treatment)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>QoY 12</td>
<td>1.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 13</td>
<td>0.92</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>QoY 14</td>
<td>1.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Observations</td>
<td>95001</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Note: Apart from the displayed covariates, also six occupational category dummies were included.
Figure 19: Starting dates of internally provided programs

(a) CTC
(b) Vacancy referral
(c) Tests (different types)
(d) Employment on trial basis
(e) Workshop
(f) Entrepreneurship support program
Figure 20: Difference-in-differences ($\Delta^2(t)$) with 95% confidence intervals

(a) Jan 2010-Oct 2009

(b) Dec 2009-Sep 2009

(c) Nov 2009-Aug 2009

(d) Oct 2009-Jul 2009
Figure 21: Difference-in-differences in treatment share ($\Delta^2(t)$) with 95% confidence intervals

(a) Jan2010-Oct2009

(b) Dec2009-Sep2009

(c) Nov2009-Aug2009

(d) Oct2009-Jul2009
Figure 22: Dif-in-dif average effect with 95% confidence intervals

(a) Jan2010-Oct2009
(b) Dec2009-Sep2009
(c) Nov2009-Aug2009
(d) Oct2009-Jul2009

Figure 23: Estimated propensity scores for treated and non-treated individuals
Figure 24: Static matching estimators