

# Comparing Econometric Methods to Empirically Evaluate Activation Programs for Job Seekers

Paul Muller\*      Bas van der Klaauw\*\*      Arjan Heyma\*\*\*

February 12, 2020

## Abstract

We test whether different identification strategies give similar results when evaluating activation programs. Budgetary problems at the Dutch unemployment insurance (UI) administration in March 2010 caused a sharp drop in the availability of these programs. Using administrative data provided by the UI administration, we evaluate the effect of the program (1) exploiting the policy discontinuity as a quasi-experiment, (2) using dynamic matching assuming conditional independence, and (3) applying the timing-of-events model. All three strategies use the same data to consider the same program in the same setting, and show that the program reduces job finding directly after enrollment. However, the magnitude of the estimated drop in job finding differs between the three estimation methods. In the longer run, all three methods show a zero effect on employment.

Keywords: policy evaluation, active labor market programs, unemployment duration

JEL-code: J64, C52, C41

## 1 Introduction

In 2002, the Dutch market for activation programs was privatized, implying that the unemployment insurance (UI) administration buys services of private companies

---

\*VU University Amsterdam

\*\*VU University Amsterdam and Tinbergen Institute

\*\*\*SEO Economic Research

VU University Amsterdam, Department of Economics, De Boelelaan 1105, 1081 HV Amsterdam. Email: p.muller@vu.nl, b.vander.klaauw@vu.nl, a.heyma@seo.nl.

This paper has previously circulated under the title "*Comparing Econometric Methods to Empirically Evaluate Job Search Assistance*". We thank UWV, and in particular Han van der Heul for making the data available and providing information on institutions. This paper has benefited from comments and suggestions by Grégory Jolivet, Jaap Abbring, Mike Elsby, Bart Cockx, Bastian Ravesteijn, three anonymous reviewers and seminar participants at the EALE (2013), ESPE (2014), VU University Amsterdam, University of Edinburgh, BeNA Berlin, University of Vienna, University of Copenhagen, University of Gothenburg, DIW Berlin, IZA Bonn and Collegio Carlo Alberto Turin. Van der Klaauw acknowledges financial support from a Vici-grant from the Dutch Science Foundation (NWO).

to assist benefits recipients in their job search. Due to the economic crisis, the demand for programs increased sharply in 2009 and early 2010, leading to budgetary problems in March 2010. The government refused to extend the budget, which terminated the purchase of new programs within a period of two weeks. New UI benefits recipients could no longer enroll in these programs. In this paper, we exploit this policy discontinuity to evaluate the effects of participating in the programs on job finding. In addition, we estimate the same effects using non-experimental estimators (matching and timing-of-events) and compare the results of the three identification strategies.

The main challenge in evaluating active labor market 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 evaluation studies use longitudinal data and compare a treatment group with a control group that is typically constructed by matching on observed characteristics (e.g. Brodaty et al. (2001), Sianesi (2004) and Lechner et al. (2011)). Some studies exploit institutional features or policy changes that generate random variation in program participation (e.g. Dolton and O’Neill (2002) and Cockx and Dejemeppe (2012)). Less than 10% of the studies use an experimental design (e.g. Van den Berg and Van der Klaauw (2006), Graversen and Van Ours (2008) and Card et al. (2011)). While Card et al. (2010) and Kluve (2010) find no evidence of a relationship between the identification strategy and the empirical results in their meta-analyses, we know since LaLonde (1986) that non-experimental estimators may produce results that do not concur with those from experimental evidence. Heckman et al. (1997) stress that experimental and non-experimental estimates can only produce similar results if three requirements are fulfilled. First, the data source should be the same for the treatment and control group. Second, treated and control individuals should be active in the same local labor market. Third, the data should contain a rich set of variables that affect both program participation and labor market outcomes. Smith and Todd (2005) argue that each of these requirements is likely to be violated in the non-experimental evaluations by LaLonde (1986).<sup>1</sup> However, Bléhaut and Rathelot (2014) show for a French job search assistance program that even when all conditions

---

<sup>1</sup>Dehejia and Wahba (1999) find that matching estimators produce results closer to the experimental evidence using the same data as LaLonde (1986), but their results are disputed by Smith and Todd (2005).

are satisfied the estimated average treatment effect on the treated is very different when using a random control group compared to using a matching estimator on the usual non-participants in the treatment group.<sup>2</sup>

We contribute to this literature by empirically comparing identification strategies. Our contribution is threefold. First, our quasi-experimental estimates are identified from a nationwide policy discontinuity in 2010 and we use high-quality administrative data.<sup>3</sup> Therefore we observe, for a large sample of individuals around the policy discontinuity, detailed information on individual characteristics, pre-unemployment labor market variables, current unemployment spell characteristics and the timing of privately provided activation programs. Second, for the non-experimental analysis we use a recently proposed dynamic matching estimator (Vikström (2017)) as well as the timing-of-events model (Abbring and Van den Berg (2003)). All our identification strategies take the dynamic nature of program assignment into account, and correct for dynamic selection.<sup>4</sup> We discuss our identification strategies within a framework for treatment effects with dynamic enrollment (see Kastoryano and Van der Klaauw (2011) for a more extensive discussion). Third, we consider the *private* provision of activation programs that are commonly offered by *public* institutions. Current evidence on the effectiveness of private providers varies: Krug and Stephan (2016) show that placement services are less effective when offered privately and Behaghel et al. (2014) and Cottier et al. (2015) find that assistance provided through the private market is ineffective in stimulating exit to work. On the other hand, Bennemarker et al. (2013) find positive effects of privately provided employment services.

The policy discontinuity that we consider reduced the weekly enrollment in the privately provided activation programs from 1300 to less than 80 within one month and it remained below 50 afterwards. We estimate the treatment effect on the treated by comparing employment outcomes of cohorts entering unemployment relatively shortly after each other. Since these cohorts reach the policy discontinuity

---

<sup>2</sup>Mueser et al. (2007), Lalive et al. (2008), Biewen et al. (2014) and Kastoryano and Van der Klaauw (2011) provide recent comparisons of (quasi-)experimental and non-experimental estimators for active labor market programs.

<sup>3</sup>We refer to the program discontinuity as a “quasi-experiment”.

<sup>4</sup>Evaluations with dynamic enrollment struggle to find the appropriate control group (Fredriksson and Johansson (2008)). Often the dynamic setting is simplified to static enrollment (Brodsky et al. (2001), Bléhaut and Rathelot (2014) and Sianesi (2004)).

at different unemployment durations they are affected differently, which identifies the effect of the programs (Van den Berg et al. (2014)). Seasonal variation in the labor market is controlled for using cohorts from the previous year. The estimation results show that after starting a program, employment probabilities are reduced significantly. After two to three months these increase again, up to a zero difference in employment 12 to 18 months after enrolling in the program.

Where the quasi-experimental estimator compares employment outcomes between different cohorts, the matching estimator and timing-of-events model compare treated and non-treated individuals within the same cohorts. The timing-of-events model allows for selection on unobservables, but adds more structure to the model. The estimation results of both approaches show again a significantly negative effect of program participation directly after entering the program, which disappears in the longer run. These results are in line with our quasi-experimental estimates, but the magnitude of the negative effect varies across the different estimators and differences are sometimes significant.<sup>5</sup> The dynamic matching estimator and the timing-of-events model can use larger samples than the quasi-experimental approach. Therefore, the quasi-experimental treatment effects are estimated much less precise, also because intention-to-treat effects are inflated with small shares of program participants.<sup>6</sup> The results from the dynamic matching estimator and the timing-of-events model are both robust against extending the sample beyond the discontinuity cohorts, but the treatment effects estimated using the timing-of-events model are closer to zero. This may suggest that there is some selection on unobservables and in particular more disadvantaged unemployed workers are more likely to participate in the program.

The results from the three different approaches yield the same policy recommendations, but interpreting results should be done with care. The quasi-experimental results rely on a common trend assumption, stating that in the absence of the policy discontinuity, different cohorts experience similar employment probabilities, up to a constant difference. Furthermore, the activation programs contain a mixture of caseworker meetings, job referrals, training or schooling, subsidized employment

---

<sup>5</sup>The negative impact on employment in the short run is in line with the well-documented lock-in effect (Lechner et al. (2011)).

<sup>6</sup>In absolute terms the monthly number of individuals entering the program is large, but this remains a small fraction of all eligible UI benefits recipients in the Netherlands.

and goal setting in the job search process. These elements are common in activation programs in other countries, but with different intensities. The programs are offered in addition to the “basic” assistance of the UI administration (mostly irregular meetings with caseworkers). Caseworkers have substantial discretion when assigning job seekers to programs, which is common to many UI administrations. Finally, our observation period describes the period of recession during the economic crisis and relates to the Dutch setting, which offers similar benefits levels as other Northern-European countries, but more generous benefits than the US or UK.

The remainder of the paper is structured as follows. We describe the institutional setting and the budgetary problems which led to the policy discontinuity in Section 2. An overview of the data is provided in Section 3. In Section 4 we define our treatment effect of interest. In Section 5 we present non-experimental results from the matching and timing-of-events estimators. In Section 6 we discuss how the discontinuity allows identification of the treatment effect and present estimation results. Section 7 compares the results from the different approaches and provides a discussion. Section 8 concludes. Supplementary material and more results are available in the Online Appendix.

## **2 Institutional setting and the policy discontinuity**

In the Netherlands, UI is organized at the nationwide level. During our observation period, the UI administration (UWV) paid benefits to workers that involuntarily lost at least five working hours per week (or half of their working hours if this is less than five). Workers should have worked for at least 26 weeks out of the last 36 weeks. Fulfillment of this “weeks condition” provides eligibility to benefits for three months. If the worker has worked at least four out of the last five years, the benefits eligibility period is extended by one month for each additional year of employment. The maximum UI benefits duration is 38 months. During the first two months benefits are 75% of the previous wage, capped at a daily maximum. From the third month onward it is 70% of the previous wage (see De Groot and Van der Klaauw (2019) for a more extensive discussion).

A UI benefits recipient is required to search for work actively, which implies

making at least one job application each week. Benefits recipients are obliged to accept any suitable job offer.<sup>7</sup> Caseworkers at the UI administration provide basic job search assistance and monitor compliance to job search obligations, but the intensity of individual meetings is low. In 2009, if caseworkers judged that the benefits recipient required more than the usual guidance, they could assign an individual to a program with the goal to increase the job finding rate. 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 UI administration, while others were purchased externally from private companies. These were for-profit companies and they faced payment schemes that consisted partly of lump-sum payments and partly of performance payments.<sup>8</sup>

Our analysis focuses solely on the programs that were externally provided by private companies. These can be broadly classified as (with relative frequency in parentheses) job search assistance programs (50%), training or schooling (28%), subsidized employment and other programs (22%). However, different programs were frequently combined, so schooling is often followed by job search assistance or job search assistance can be accompanied with training or followed by subsidized employment. We consider the first moment of participating in a program as the start of treatment, which implies that our analysis relates to the mixture of programs. Though some guidelines existed, caseworkers had a large degree of discretion in deciding about the assignment of the different programs.

The lack of centralized program assignment together with an increased inflow in unemployment due to the recession caused that many more individuals were assigned to these programs in 2009 and early 2010 than the budget allowed. Therefore, the entire budget had been exhausted by March 2010. Authorities refused to extend the budget and declared that no new programs could be purchased from that moment onward.<sup>9</sup> Assistance offered internally by the UI administration continued without

---

<sup>7</sup>During the first six months of unemployment a suitable job is defined as a job at the same level as the previous job, between six and 12 months it can be a job below this level, and after 12 months any job is suitable and should be accepted.

<sup>8</sup>The performance payments were not very sophisticated and varied between contracts. Often an additional payment would be made if a certain share of the program participants would leave the UI benefits system within a particular time period.

<sup>9</sup>This was declared by the Minister of Social Affairs, in a letter to parliament on March 15. An exception was made for a small number of specially targeted programs (mostly for long-term

change. In Section 3 we show that indeed the number of new program entrants dropped to almost zero in March 2010 and remained very low afterwards.

In our empirical analysis, we focus on job finding. Van der Klaauw and Ziegler (2019) state that about 30% of all exits from UI are towards jobs provided by temporary employment agencies. There the first job contract is typically around six months, but working hours are sometimes uncertain. Also benefits recipients who find regular jobs start often with a temporary contract of about six months to one year. De Groot and Van der Klaauw (2019) and Van den Berg and Van der Klaauw (2019) report that wages in the first job are, on average, about 7 to 9% higher than the benefits level.

### 3 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 data contain 583,193 UI spells (for some individuals there are multiple spells).<sup>10</sup> 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 benefits payments. Right censoring occurs on January 1st, 2012, when our data were constructed, so for each spell we can observe at least 16 months of potential benefits receipt. The data contain a detailed description of all activation programs (both internally and externally provided) in which benefits recipients participated. Furthermore, individual characteristics and pre-unemployment labor market outcomes are included.

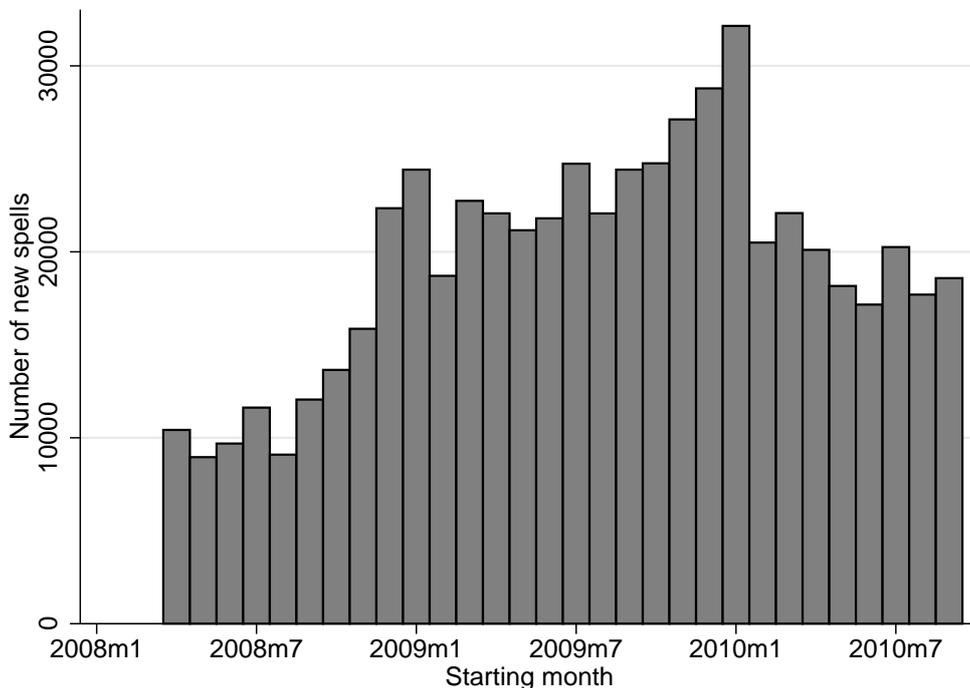
Figure 1 shows how the monthly number of individuals entering UI evolves over time. The economic crisis caused a substantial increase in the inflow per month from around 10,000 in 2008 to 30,000 by the end of 2009. From 2010 onward the inflow decreased somewhat. Table 1 presents summary statistics for the full sample, as well as for three cohorts defined by their month of inflow into unemployment.

---

unemployed workers).

<sup>10</sup>The original dataset contains 671,743 spells. We exclude 34,968 spells from individuals previously employed in the public sector and 17,454 from individuals older than 60 years. Next, we drop 25,778 spells for which important variables are missing, 524 spells from individuals working more than 60 hours or less than 12 hours in their previous job and 9,504 spells during which the job seeker also had ongoing employment. Finally, we exclude 288 spells with a negative unemployment duration and 34 spells with negative benefits eligibility.

Figure 1: Number of UI entrants per month



Column (1) shows that for the full sample the median duration of unemployment is 147 days (almost five months). About 44% of those exiting UI find work and for 13% the reason of exit is unknown. We include both in our outcome measure “outflow to work”.<sup>11</sup> The end of the benefits entitlement period is reached for 32% of the spells. Almost 5% leave unemployment due to sickness or disability; the rest leave for other reasons.

In the full sample about 13% of the benefits recipients participate in one of the externally provided programs. Slightly more than half of these programs focus on job search assistance, over a quarter involve some sort of training, and less than a quarter consist of some other program.<sup>12</sup> Note that these numbers refer only to the *first* program that an individual participates in. About 26% of all individuals

<sup>11</sup>Often UI recipients do not give a reason for terminating their benefits or these reasons are not registered by caseworkers. In most cases the individual starts generating income from other sources related to working. Our results are robust against including unknown as outflow to work.

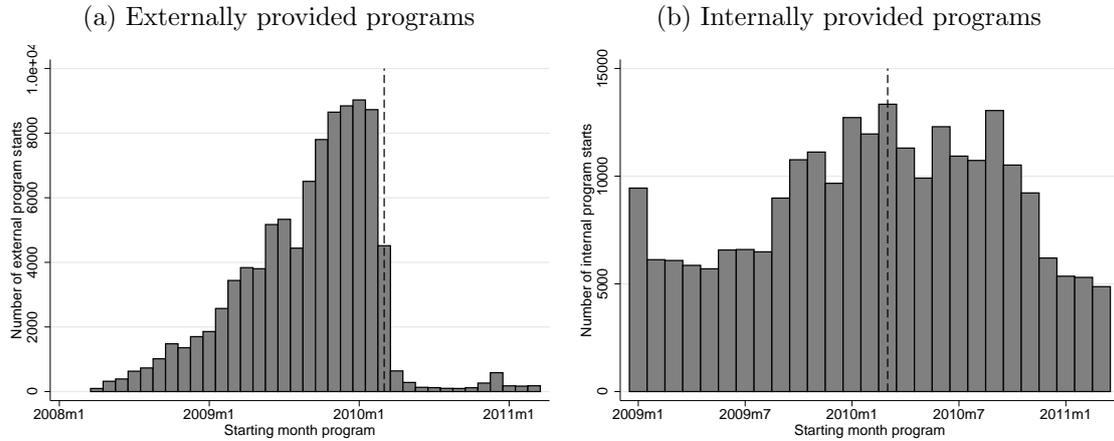
<sup>12</sup>To increase precision in exploiting the policy discontinuity, it is possible to select individuals with characteristics that are correlated to participation in the external activation programs. This has been done in the Discussion Paper version of this paper (Muller et al. (2017)). The gains of reducing the sample are small though, because the correlation between individual characteristics and program participation is weak.

Table 1: Descriptive statistics

	Inflow cohort UI:			
	Full sample	April 2008	April 2009	April 2010
Unemployment duration (median, days)	147	123	161	158
Reason for exit (%):				
Work	43.8	45.1	40.6	46.3
Unknown	12.7	13.0	12.7	12.7
End of entitlement period	31.9	27.7	35.4	28.2
Sickness/Disability	4.6	4.9	4.7	5.1
Other	7.0	9.3	6.7	7.8
Participation external program (%):				
Any program	13.3	16.1	22.3	0.2
Job search assistance <sup>a</sup>	6.8	9.8	12.8	0.1
Training <sup>b</sup>	3.6	4.0	5.4	0.0
Other	2.9	2.4	4.2	0.1
Participation internal program (%):				
Any program	26.1	6.1 <sup>c</sup>	28.9	30.4
Tests	7.1	0.8 <sup>c</sup>	8.6	8.2
Subsidized employment	1.1	0.3 <sup>c</sup>	1.0	1.5
Workshops entrepreneurship	2.0	0.8 <sup>c</sup>	2.6	1.8
Job search assistance	5.3	0.6 <sup>c</sup>	4.7	8.3
Other	12.7	4.4 <sup>c</sup>	14.6	12.4
Gender (% female)	41.5	50.8	42.0	44.0
Age	37.9	38.8	37.7	38.3
Immigrant (%)	8.8	10.0	8.7	8.3
Education (%):				
Low	30.8	33.0	29.5	28.7
Middle	46.9	44.1	46.7	46.7
High	22.3	22.9	23.7	24.6
UI history last 3 year (%)	31.6	38.7	28.3	28.3
Previous income (%):				
Low	20.4	26.1	20.5	18.9
Middle	63.1	58.0	63.4	62.7
High	16.4	15.9	16.1	18.4
Unemployment size (hours) <sup>d</sup>	33.7	32.9	34.0	33.4
Observations	583,193	10,422	22,069	20,107

<sup>a</sup> Job search assistance contains ‘IRO’ (Individual reintegration agreement), ‘Job hunting’ and ‘Application letter’. <sup>b</sup> Training contains ‘Short Training’ and ‘Schooling’. <sup>c</sup> Biased downwards, because participation in internal programs was rarely recorded before 2009. <sup>d</sup> Unemployment size is the number of hours per week for which an individual is unemployed, based on the number of weekly working hours in the last job.

Figure 2: Monthly number of UI benefits recipients entering an activation program (the dashed lines indicate the policy discontinuity in March 2010)



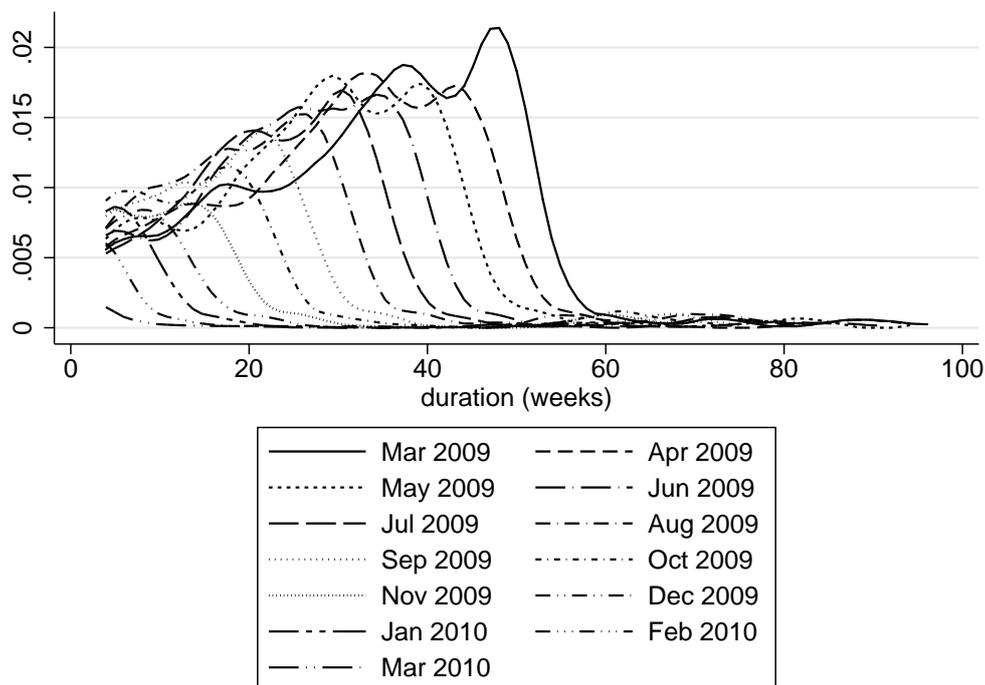
participate in an internal program, of which the majority is either some test (such as a competencies test) or job search assistance.

The dataset contains many individual characteristics. The lower panel of Table 1 presents sample means for some characteristics. The average individual is almost full-time unemployed (34 hours) and 32% have experienced an earlier period of unemployment in the three years prior to entering UI. In columns (2)-(4) the same statistics are presented for three subgroups of individuals entering unemployment in April 2008, April 2009 and April 2010, respectively. The impact of the policy discontinuity in March 2010 becomes clear, as the share that participate in an external program drops to almost zero in the April 2010 cohort. To illustrate the impact of the discontinuity in March 2010, we show the number of external programs started per month in panel (a) of Figure 2. The dashed line indicates the policy change in March 2010.<sup>13</sup>

Each inflow cohort reaches the policy discontinuity at a different moment during their UI spell. Figure 3 illustrates that each subsequent cohort experiences the drop in the first program entry hazard one month earlier in their unemployment duration. Participation in an external program is not restricted to a certain unemployment duration. Before the policy discontinuity the participation hazards of the different cohorts are very similar, indicating that there were no earlier policy changes. The cohort of March 2010 has an almost zero probability of entering an external program.

<sup>13</sup>Figure D1 in Online Appendix D shows that the discontinuity occurs for all types of programs.

Figure 3: Hazard rate into the external programs by month of inflow

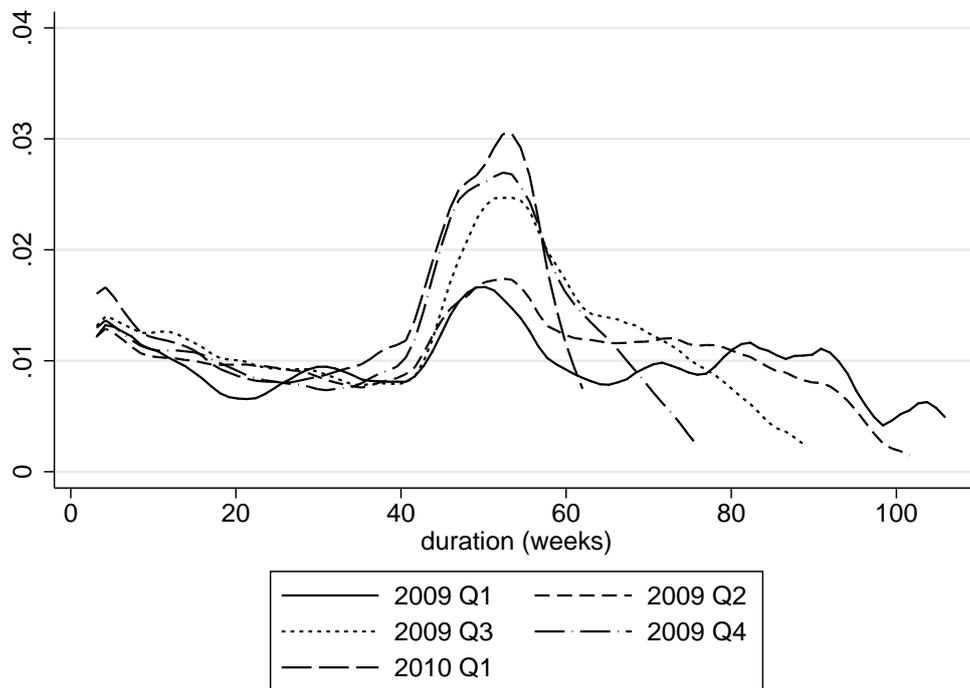


A concern might be that caseworkers have responded to the inability to assign unemployed workers to external programs. However, resources for the internal programs (offered by the UI administration itself) remained unaffected around March 2010, limiting the scope for scaling up internal programs. The number of internal programs starting in each month is shown in panel (b) of Figure 2. There is no indication of a response around the date of the policy discontinuity. Separate graphs by type of program are provided in Online Appendix D (Figure D3). The hazard rates into an internal program for different cohorts are shown in Figure 4. The hazard rates are very similar, supporting the assumption that internal program provision was unaffected by the policy change.<sup>14</sup>

A further concern might be that even though the number of internal programs was not changed, caseworkers 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

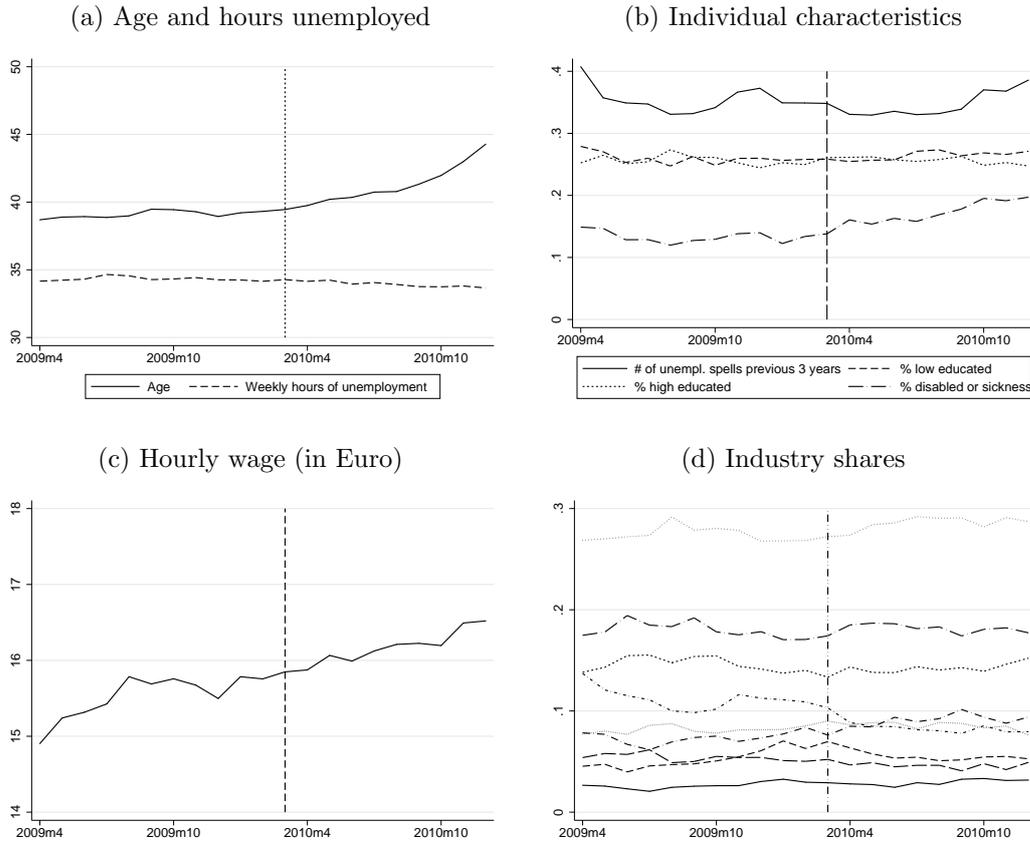
<sup>14</sup>In theory, job seekers could decide to pay for an external program themselves once it is no longer offered through the UI administration. However, the costs of these programs are considerable, especially for unemployed workers such that this almost never happens in the Netherlands.

Figure 4: Hazard rate into the internal programs by quarter of inflow



change external program participation to no participation, but, for some individuals, changes it to internal program participation. To investigate whether such a shift of internal program targeting indeed occurred, we present sample means of characteristics of individuals enrolling in an internal program per month in Figure 5. Mean age and weekly hours of unemployment are shown in panel (a) of Figure 5, unemployment, disability history and education level are shown in panel (b), the previous hourly wage is shown in panel (c) and the share of nine industry categories is shown in panel (d). None of the graphs indicate any kind of discontinuity around March 2010, which suggests that caseworkers did not shift internal programs to individuals who would otherwise enroll in an external program. The effect of enrollment in an external program can thus be interpreted as the effect conditional on the allocation of internal programs.

Figure 5: Composition of the internal program participants



## 4 Treatment effects

In this section we define the treatment effects that we aim to estimate. Recall that only a small share of all unemployed workers enter an external program during their unemployment spell. Due to selectivity in the participation decision, the composition of program participants and non-participants is different. We focus on treatment effects for participants. These treatment effects are nonstandard because enrollment in the program is dynamic.

Our outcome variable is duration until employment, which is a random variable denoted by  $T > 0$ . Define  $Y_t = \mathbf{1}(T < t)$ , a variable equal to one if the individual finds work within  $t$  time periods, and zero if the individual is still unemployed. Define  $S$  to be a random variable denoting the duration at which program participation starts. Potential unemployment durations  $T^*(s)$  depend on program participation  $S = s$ . Since program assignment is dynamic, even for a single type of program many different treatment effects arise. The program can start at different durations, while

the effect can be measured at different points in time (see for an extensive discussion of dynamic treatment effects Abbring and Heckman (2007)). We define potential outcomes when treated as:

$$Y_{1,t}^*(s) = \begin{cases} 1 & \text{if } T^*(s) < t \\ 0 & \text{if } T^*(s) \geq t \end{cases}$$

The potential outcome under no treatment is defined as  $Y_{0,t}^* = \lim_{s \rightarrow \infty} Y_{1,t}^*$ . We adopt the so-called no-anticipation assumption (Abbring and Van den Berg (2003)), which imposes that program participation at duration  $s$  only affects potential outcomes at durations  $t > s$ . This is required to define counterfactuals and thereby treatment effects. The no-anticipation assumption allows us to write the potential untreated outcomes as

$$Y_{0,t}^* = Y_{1,t}^*(s) \quad \forall s > t$$

The no-anticipation assumption is strict since it rules out that individuals change their job search behavior prior to  $s$  in anticipation to learning that they will enroll in the program at time  $s$ .<sup>15</sup> Such behavior may be unlikely for the external programs we consider in this paper. Programs are assigned by caseworkers on an individual basis. There are no strict criteria for participation and only a small fraction of the unemployed workers can enroll, so it is impossible for job seekers to know in advance when they will enter a program.

Individuals leave unemployment after different durations, such that the composition of the survivors changes over time. This dynamic selection requires defining the subgroup for which the treatment effect is evaluated (see Van den Berg et al. (2014)). We are interested in the treatment effect on the treated, which is the effect of program participation on individuals treated at duration  $s$ . Since individuals exit from unemployment at any duration, we condition on remaining unemployed until at least duration  $s$ . This (conditional) contrast between the potential outcome under treatment at  $s$  and no treatment is defined by Van den Berg et al. (2014) as

---

<sup>15</sup>The no-anticipation assumption does not rule out that treatment assignment probabilities differ between individuals and that individuals are aware of this. It also allows that program enrollment is more likely in some periods than in other periods, which may be known by job seekers. Only the exact timing of the program start should be unanticipated.

the average treatment effect on the treated survivors ( $\text{ATTS}(s, t)$ ):

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

The average treatment effect on the treated survivors  $\text{ATTS}(s, t)$  provides a series of effects for different values of  $s$  and  $t$ . Policy makers often focus on all participants, which requires knowledge about the treatment assignment policy. Let  $f(s)$  denote the density function of starting treatment at time  $s$  and  $S_0(s)$  the survivor function in unemployment until time  $s$  for individuals who did not participate in treatment before  $s$ . We define the average treatment effect on the treated evaluated at time  $t$  after entering unemployment as

$$\text{ATEET}(t) = \frac{\int_{s=0}^t f(s)S_0(s)\text{ATTS}(s, t)ds}{\int_{s=0}^t f(s)S_0(s)ds} \quad (2)$$

This describes the average treatment effect on those individuals who started participating in the treatment before time  $t$ . For tractability of the comparison of the different estimation approaches, we focus on estimating the  $\text{ATTS}(s, t)$  and only briefly discuss the estimated  $\text{ATEET}(t)$  in Section 7.

## 4.1 Choice of samples

The different identification strategies that we apply, require sample selections that do not necessarily coincide. Exploiting the policy discontinuity requires using a specific sample of individuals entering unemployment around the time of the discontinuity. Matching and timing-of-events can use a much larger sample including individuals entering unemployment earlier or later. To apply each identification strategy as they would be applied in a usual application, we construct two subsamples from the full sample (see Table 2).

A difference in results can be due to different sample selection rules. We argue that choice of sample is an essential part of an identification strategy. However, to investigate to what extent the sample choice causes differences in results, we perform each analysis also with a smaller sample that is the same across all strategies (column (2) in Table 2). A more extensive discussion on the selection of the smaller sample is presented in Section 6, where we discuss the quasi-experimental approach exploiting the policy discontinuity. In addition, we apply the matching and timing-of-events

Table 2: Identification strategies and samples

	(1)	(2)	(3)
	Full sample: Inflow between April '08 and Sept. '10	Discontinuity sample: Inflow between Oct. '09 and Jan. '10	Pre-disc. sample <sup>a</sup> Inflow between April '08 and Jan. '10
Matching	yes	yes	yes
Timing-of-events	yes	yes	yes
Quasi experiment	no	yes	no
Observations	583,193	112,678	428,160

<sup>a</sup> In addition to restricting the inflow period, observations in this sample are also censored at the discontinuity.

estimators to a third sample that excludes the discontinuity period (column (3) in Table 2). This sample contains only individuals entering unemployment before the discontinuity and censors all observations at the time of the discontinuity. The rationale for analyzing such a sample is that the discontinuity creates exogenous variation in program participation, and we study how non-experimental methods perform without including such variation.

Our comparison considers the approaches presented in Table 2. The full sample and the pre-discontinuity sample are used for the non-experimental methods only, while the discontinuity sample is used for all three methods.

## 5 Non-experimental analysis

### 5.1 Matching estimator

We start the empirical analysis by applying a matching estimator. The identification strategy does not exploit the policy discontinuity, but instead compares individuals with similar characteristics differing only in treatment status. We apply a dynamic matching estimator (proposed by Vikström (2017)) to account for the dynamic setting and dynamic selection.

Matching methods construct counterfactual outcomes in the absence of randomized assignment (e.g. Imbens and Wooldridge (2009) and Heckman et al. (1997)). The approach relies on two main assumptions. First, selection into treatment is on

observables only:

$$Y_{0,t}^*(s), Y_{1,t}^*(s) \perp S | X = x \quad (3)$$

This unconfoundedness assumption implies that after conditioning on a set of observed characteristics, assignment to treatment is independent of the potential outcomes. See Vikström (2017) for a discussion of how this assumption generalizes to a setting of dynamic treatment assignment. Our administrative data include a rich set of covariates, which is crucial for the likelihood that unconfoundedness holds. Employment histories are argued to be particularly important, because they tend to be strong predictors of future labor market outcomes as well as program participation (e.g. Card and Sullivan (1988), Heckman et al. (1999), Gerfin and Lechner (2002), Lechner et al. (2011) and Lechner and Wunsch (2013)). In addition to employment history (previous hourly wage, unemployment history, industry), we observe individual characteristics (age, gender, education level, marital status, region) and variables describing the current unemployment spell (unemployment size in hours, sickness or disability, maximum benefits entitlement). This set of covariates is at least as extensive as usually available when evaluating active labor market programs.

Second, the matching estimator requires a common support in the distribution of the covariates between program participants and nonparticipants. For our dynamic setting we assume

$$f(s; x) > 0 \quad \forall x, s$$

where  $f(s; x)$  is the density function of enrolling in the program after  $s$  periods of unemployment conditional on the set of covariates  $x$ . At any duration, all individuals have a positive probability of starting treatment regardless of their characteristics. This ensures that if the sample size is sufficiently large, counterfactuals can always be found. This assumption is likely to hold, since there are no (combinations of) individual characteristics that perfectly predict program participation in our data.

As baseline, we consider the full sample, and as robustness checks we use two subsamples (as discussed in Subsection 4.1). We compare those that are treated at  $s$  with those that are never treated. Clearly, the probability of never receiving treatment depends on the length of an unemployment spell. That is, the longer someone remains unemployed, the larger the probability of starting treatment. To correct

for the dynamic selection, we apply the estimator proposed by Vikström (2017).<sup>16</sup> The treatment group contains all individuals that start treatment at  $s$ , while the control group includes all individuals that have not been treated and are still unemployed by duration  $s$ . When measuring the outcome at time  $t > s$ , only those in the control group that have not yet started treatment at time  $t$  are included (spells of those in the control group that start treatment between  $s$  and  $t$  are censored). Since program assignment is not random, the surviving control group individuals are weighted using propensity scores, making them comparable to the treatment group in terms of observed characteristics. Given the continuously changing composition of the control group, the weights are recalculated for each period. The estimator thus corrects for both selective exits and potentially selective right-censoring. A formal description of the estimator is provided in Online Appendix A.

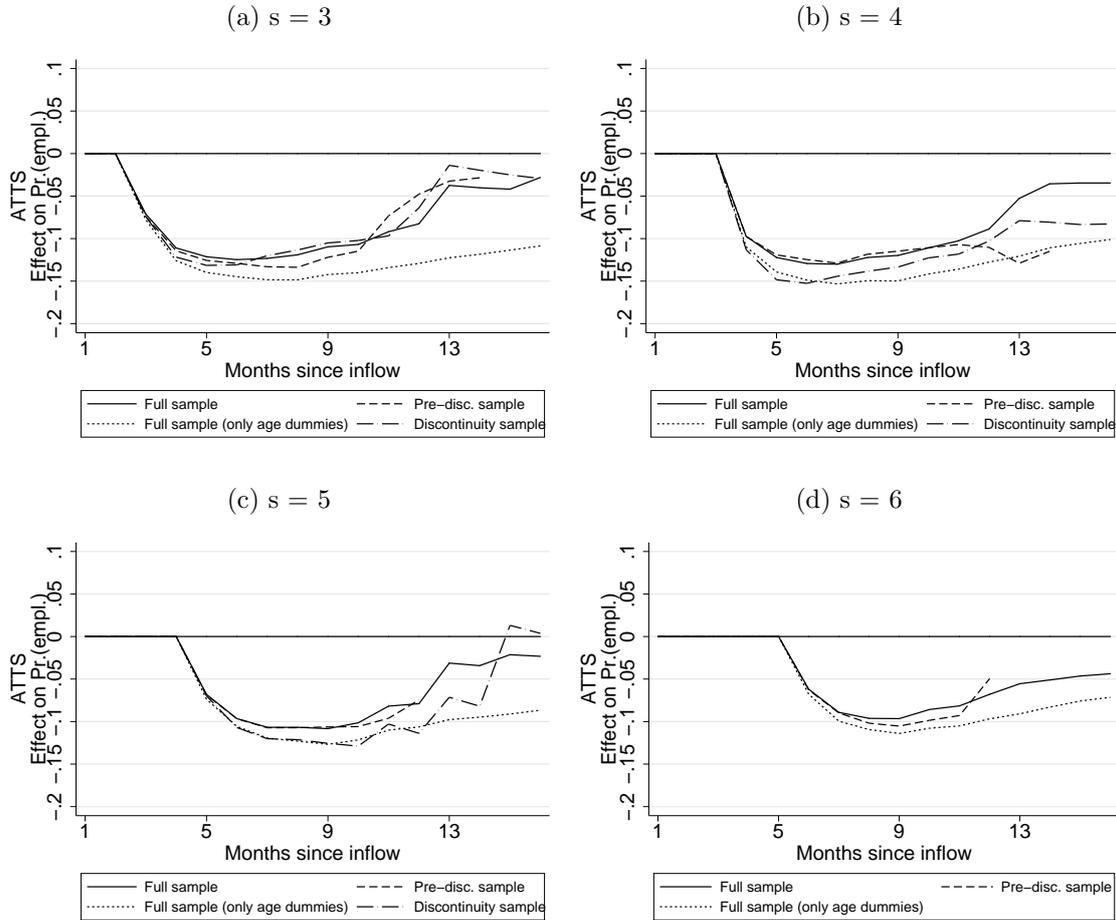
We aggregate all spell data into 30-day intervals and estimate the impact of treatment on job finding for different values of  $s$ , that is, programs starting at different elapsed durations. The estimates are presented in Figure 6 for treatment starting at  $s = 3, 4, 5, 6$  months, and measured up to an unemployment duration of 16 months. We focus on these four values of  $s$  because inflow into programs is highest at these durations (see Figure D2 in Online Appendix D). The presented estimates are differences in survivor probabilities between the treatment group and the (weighted) control group.

The estimation results show that program participants are less likely to find employment. Around two to three months after the program has started, the probability of finding employment in the control group is 10 to 20%-points higher than in the treatment group. This difference remains for at least 12 months. The pattern of program effects is very similar across the four panels (a)-(d) and estimated impacts for other values of  $s$  ( $s = 1, \dots, 8$ ) show very similar results (see Figure D4 in Online Appendix D). The effects of the program are thus equally negative when a benefits recipient enters the program relatively fast or later during the period of unemployment. The declining impact of program participation during the first couple of months is consistent with a lock-in effect. Recall that we estimate the effect of a mixture of activation programs that regularly includes schooling or training.

---

<sup>16</sup>An alternative approach that has been applied in the literature is to compare those treated at  $s$  with those not *yet* treated at  $s$  and interpret the difference as the effect of early treatment relative to potential later treatment (see Sianesi (2004)).

Figure 6: Dynamic matching ATTS estimates: matching estimates based on different samples



Note: Estimates based on too few observations are not shown in the graphs. In panel (d) the discontinuity sample is omitted, because no individual starts a program beyond six months of unemployment. Figures D4 - D7 in Online Appendix D show separate graph for all estimates also including confidence intervals. The negative effects are statistically significant until at least 12 month of unemployment. After that some estimates become insignificant, both because estimated effects move to zero and sample sizes become small.

Participation in such programs often causes that unemployed workers invest less time in job search. Furthermore, job search assistance programs may crowd out certain types of job search effort which are more effective for finding work quickly (e.g. Van den Berg and Van der Klaauw (2006)).

Figure 6 shows that using different samples yields very similar estimates for the impact of program participation. In the discontinuity sample selectivity in program participation should be less of an issue than in the other samples. The similarity of estimated effects can either imply that the set of covariates is sufficiently rich to deal with selective program assignment or that programs are assigned relatively

random. Therefore, we repeat the matching estimator, but only using age dummies as observed covariate. This hardly affects the estimates, which provides evidence that selectivity on observed covariates is not large when assigning benefits recipients to the programs. Whether selectivity on unobserved characteristics is important when assigning the program requires comparing the findings to those from the other empirical approaches.

We conclude that the negative impact of the program on employment probabilities is (1) the same for assignment at different durations, (2) very robust across different samples and (3) even robust against excluding most observed characteristics from the matching estimator.<sup>17</sup>

## 5.2 Timing of events model

Matching requires very modest functional-form assumptions, but relies on a potentially strong unconfoundedness assumption. The timing-of-events model (Abbring and Van den Berg (2003)) allows for selection on unobserved variables, but makes stronger functional-form assumptions. This model has been applied regularly in the recent literature on dynamic treatment evaluation.<sup>18</sup>

The timing-of-events model jointly specifies job finding and entry into the program using continuous-time duration models. To control for unobserved characteristics, the unobserved heterogeneity terms in both hazard rates are allowed to be correlated. Identification relies on the mixed proportional structure of both hazard rates. Since the model is a continuous-time model, we use daily spell data and do not have to make an assumption on which unit of time to take when discretizing the unemployment durations.

We present a concise description of the model here, while a detailed version is presented in Online Appendix B. Consider an individual entering unemployment at calendar date  $\tau_0$ . The job finding (hazard) rate depends on the number of days of unemployment  $t$  (modeled by  $\psi_e(t)$ ), on the calendar time  $\tau_0 + t$  (modeled by

---

<sup>17</sup>Recall that we also include exits from UI for unknown reasons as job finding. Figure D8 in Online Appendix D shows that estimated program effects are slightly more positive if we consider exits for unknown reasons as job finding and slightly more negative when we consider all exits as job finding. Our estimation results are thus not specific to the definition of job finding and this also holds for the estimation results in the following sections.

<sup>18</sup>See for example Abbring et al. (2005), Van den Berg et al. (2004), Lalive et al. (2005) and Van der Klaauw and Van Ours (2013). See Lombardi et al. (2018) for a more extensive list of papers that apply the model for other purposes than evaluating active labor market policies.

Table 3: Hazard ratio estimates of the treatment parameters in the timing-of-events model

	(1)		(2)		(3)	
	Full sample		Discontinuity sample		Pre-discont. sample	
	haz.ratio	st.er.	haz.ratio	st.er.	haz.ratio	st.er.
Program effect on UI outflow:						
$\delta_{(1-3 \text{ months})}$	0.874	0.014	0.899	0.045	0.834	0.019
$\delta_{(4-6 \text{ months})}$	1.036	0.016	0.914	0.048	0.987	0.023
$\delta_{(>6 \text{ months})}$	1.247	0.015	1.223	0.047	1.181	0.028
Observations	582,580		112,678		428,160	

$(\psi_e(\tau_0 + t))$ , observed characteristics  $x$  and unobserved characteristics  $v_e$ . When starting the activation program after  $s$  periods of unemployment, the hazard rate shifts by the treatment effect  $\delta_{t-s}$ , which can depend on the elapsed duration  $t - s$  since entering the program. The treatment effect is modeled as a piece-wise constant function of the elapsed duration since starting the program. The job finding rate is given by:

$$\theta_e(t|x, \tau_0, s, v_e) = \phi_e(t)\psi_e(\tau_0 + t) \exp \left[ x\beta_e + \delta_{t-s}I(t > s) \right] v_e \quad (4)$$

Estimation of equation (4) yields an inconsistent estimate of the treatment effects if program participation is (conditional on the observed characteristics) non-random. To account for this, program participation is modeled jointly, also using a mixed proportional hazard rate:

$$\theta_p(s|x, \tau_0, v_p) = \phi_p(s)\psi_p(\tau_0 + s) \exp(x\beta_p)v_p \quad (5)$$

with all notation similar to equation (4), but subscript  $e$  replaced by subscript  $p$ . The unobserved term  $v_p$  is allowed to be correlated with  $v_e$ , with joint discrete distribution  $g(v_e, v_p)$ . We take  $g(v_e, v_p)$  to be a bivariate discrete distribution with an unrestricted number of mass points. The duration dependence patterns and the calendar time effects are parameterized with a piece-wise constant function. Estimation of the parameters is performed by maximizing the log-likelihood, in which right-censoring is straightforwardly taken into account.

Table 3 shows the estimated hazard ratios using the full sample, the smaller discontinuity sample and the pre-discontinuity sample.<sup>19</sup> Program participation has a statistically significant negative effect on the job finding rate in the first three months ( $\delta_{(1-3 \text{ months})}$ ). In the next three months the effect is very close to zero (hazard ratio is about one). After six months ( $\delta_{(\geq 6 \text{ months})}$ ) program participation has a significantly positive effect on the probability of finding a job (1.247). The smaller discontinuity sample and pre-discontinuity sample give slightly smaller positive program effects. Standard errors are larger due to the smaller sample size. All estimation results show evidence for correlated unobserved heterogeneity, expressed by two distinct mass points.<sup>20</sup> Conditional on the observed characteristics, there is negative selection in the activation program. Unemployed workers who have unobserved characteristics that indicate longer UI spells have a higher rate of participating in the activation programs.

The estimates for the parameters  $\delta$  provide a multiplicative effect on the job finding rates. They can not be directly interpreted as measure for the treatment effects  $\text{ATTS}(s, t)$ . Therefore, we follow Kastoryano and Van der Klaauw (2011), who define for unemployed worker  $i$  with observed characteristics  $x_i$

$$E[Y_{1,t}^*(s) - Y_{0,t}^* | T > s; x_i, v_e] = \frac{\exp(-\int_0^t \theta_e(z|x_i, t, v_e) dz) - \exp(-\int_0^t \theta_e(z|x_i, s, v_e) dz)}{\exp(-\int_0^s \theta_e(z|x_i, s, v_e) dz)}$$

To translate this into the average treatment effect on the treated survivors, we should condition on the rate of receiving treatment after  $s$  periods. We use the hazard rate model for entering the program, which gives

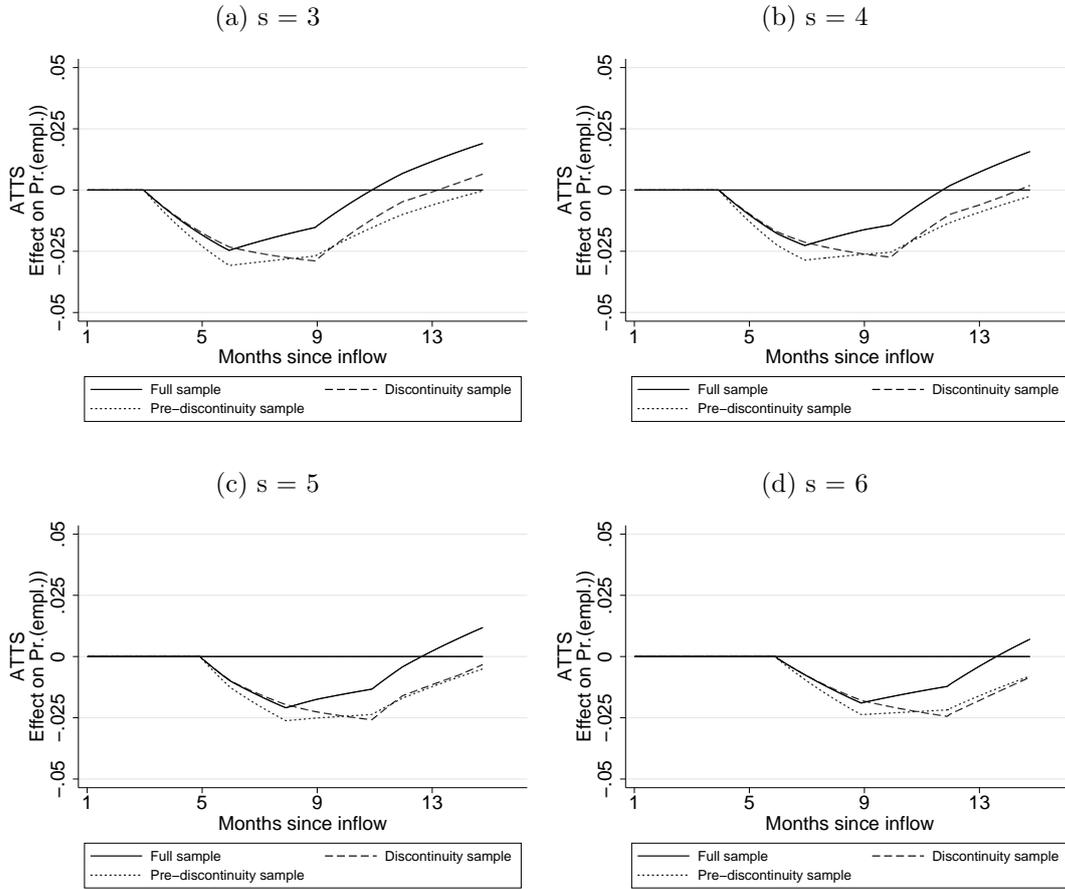
$$\text{ATTS}(s, t) = \sum_i \frac{\int_v f(s|x_i, v_p) E[Y_{1,t}^*(s) - Y_{0,t}^* | T > s; x_i, v_e] dG(v_e, v_p)}{\sum_i \int_v f(s|x_i, v_p) dG(v_e, v_p)} \quad (6)$$

where  $f(s|x_i, v_e, v_p) = \theta_p(s|x_i, v_p) \exp(-\int_0^s \theta_e(z|x_i, v_e, s) + \theta_p(z|x_i, v_p) dz)$  is the rate at which individual  $i$  enters the activation program after  $s$  periods. Since we have estimated all right-hand-side parameters in (6), we can obtain an estimate for the  $\text{ATTS}(s, t)$ .

<sup>19</sup>All estimated coefficients are presented in Table D1 in Online Appendix D.

<sup>20</sup>Note that a simulation study by Lombardi et al. (2018) finds that two mass-points are sufficient to eliminate most selection bias in a timing-of-events model.

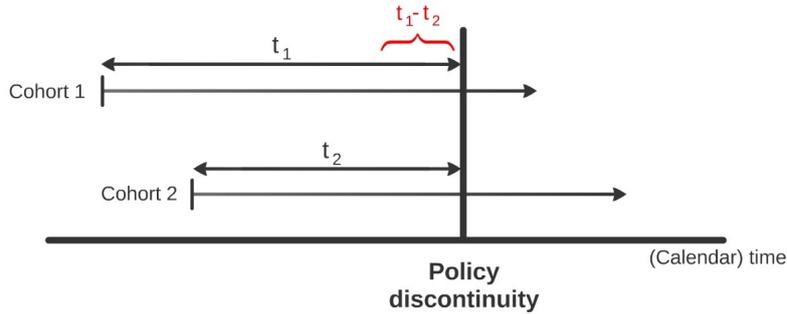
Figure 7: Dynamic matching ATTS estimates based on timing-of-events model with different samples



We present estimates for the ATTS for  $s = 3, 4, 5, 6$  in Figure 7.<sup>21</sup> Note that these are the same values of  $s$  as shown for the dynamic matching estimator in Figure 6. The graphs show that the probability of being employed decreases after starting a program, irrespective of the value of  $s$ . The magnitude of the negative impact is only slightly smaller though for later treatment ( $s = 6$ ) compared to earlier treatment ( $s = 3$ ). The different samples lead again to similar conclusions. The full sample shows somewhat more optimistic effects but the difference with the other samples is small and does not exceed the confidence intervals (see Figures D9, D10 and D11 in Online Appendix D, confidence intervals are computed using the delta-method). Finally, the overall magnitude of the negative impact in each of the four panels is smaller compared to what we found using the dynamic matching estimator.

<sup>21</sup>For computational reasons we compute the ATTS using a random subsample of 1,000 individuals.

Figure 8: Treatment effect identification



## 6 Exploiting the policy discontinuity

In this section we focus on the policy discontinuity. We first discuss how the exogenous variation due to the policy discontinuity allows to identify the effect of the program on outflow to work. Next, we discuss the estimation results.

### 6.1 Identification

The policy discontinuity creates exogenous variation in program participation over inflow cohorts in unemployment. Consider two cohorts of entrants in unemployment. The first enters unemployment  $t_1$  periods before the policy change. The second cohort enters unemployment later, but still before the policy change. For this cohort the time until the policy change equals  $t_2 < t_1$  (see Figure 8). The two cohorts face the same policy of potential program assignment for  $t_2$  time periods, implying 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 the second cohort is excluded from program participation. 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 enroll in a program prior to  $t_2$ .

To assign differences in job finding between these cohorts to program participation, three conditions should hold. First, the policy discontinuity should be unanticipated by unemployed workers and caseworkers. Job seekers and caseworkers should not have changed their behavior in anticipation to the policy change in the period just before March 2010. Our policy change has the advantage that the UI administration only realized late that the budget for these programs had run out and

expected that the Ministry of Social Affairs would extend the budget. Enrollment in the external programs stopped immediately after an extension of the budget was unexpectedly rejected.

The second condition is that there should not be compositional differences between cohorts in terms of unobserved characteristics. This condition is equivalent to a conditional independence assumption, but has milder implications. The composition of cohorts is less likely to suffer from selectivity than program assignment. We use weighting to make each cohort equivalent to the March 2010 cohort in composition of observed characteristics. We distinguish 288 groups based on interacting covariates.<sup>22</sup> Define the share of group  $g$  in cohort  $c$  by  $\alpha_{c,g}$ . The weight assigned to an observation belonging to group  $g$  in cohort  $c$  is given by:

$$w_{c,g} = \frac{\alpha_{\text{March2010},g}}{\alpha_{c,g}}$$

We define the survivor functions that will be estimated in the analysis, as the weighted average of the survivor functions of each cohort-group:

$$S_c(t) = \sum_g w_{c,g} S_{c,g}(t) \tag{7}$$

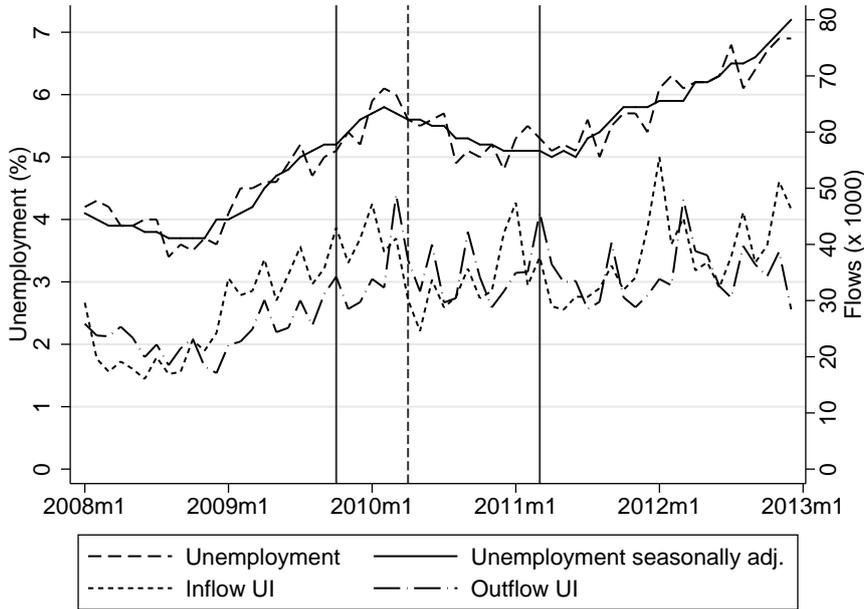
The third condition is that there should be no differences between the two cohorts in factors that affect job finding, other than the difference in program assignment. However, job finding probabilities change over the business cycle and may fluctuate due to seasonality. To reduce the impact of the business cycle and seasonality, cohorts that are close in time should be compared. But evaluating the program requires cohorts that are substantially spaced apart so that there is sufficient difference in program participation between the cohorts.

Figure 9 presents the unemployment rate and the inflow and outflow of unemployment. The vertical dashed line indicates the policy discontinuity. In the period before the policy discontinuity, 2009 and the beginning of 2010, unemployment was rising. During 2010 it decreased slightly, while in 2011 it started increasing again. In the short-run, seasonalities are the main source of fluctuations in unemployment.

---

<sup>22</sup>As characteristics we use three previous hourly wage categories, an indicator for having been unemployed in the past three years, an indicator for being married or cohabiting, age categories, three education categories and an indicator for being part-time unemployed (less than 34 hours per week).

Figure 9: Labor market indicators



Source: Statistics Netherlands, Statline.

Note: The two vertical full lines indicate the observation period that is used in the analysis.

Also inflow into and outflow from UI are relatively stable around the policy discontinuity. Changing labor market conditions may affect outcomes in two ways. First, they affect the composition of the inflow into unemployment and this affects aggregate outflow probabilities (which we address by weighting). Second, labor market conditions affect outflow probabilities directly, as it is more difficult to find work when unemployment is high (e.g. Van den Berg and Van der Klaauw (2001)).

To deal with business cycle and seasonal effects, we consider a model with an additive structure. Let cohort  $\tau$  be defined by the month of inflow into unemployment. The survivor function in unemployment ( $S(t)$ ) has some baseline shape  $\lambda(t)$  and further depends on the effect of the business cycle ( $b_\tau(t)$ ), the effect of seasonalities ( $l_\tau(t)$ ) and the effect of entering the program after  $s$  periods on being employed after  $t$  periods (which is the  $ATTS(t, s)$ ). Furthermore,  $f(s)$  describes the density function of program enrollment after  $s$  periods. Our additive specification of the survivor function is given by

$$S_\tau(t) = \lambda(t) + b_\tau(t) + l_\tau(t) + \int_{s=0}^t ATTS(s, t) f(s) ds \quad (8)$$

The policy discontinuity caused a stop on program entry at moment  $\bar{\tau}$ , without

affecting  $f(s)$  prior to this moment. Comparing the survivor function of two cohorts that entered unemployment before the policy discontinuity gives

$$S_{\tau}(t) - S_{\tau'}(t) = b_{\tau}(t) - b_{\tau'}(t) + l_{\tau}(t) - l_{\tau'}(t) + \int_{s=0}^{\bar{\tau}-\tau} \text{ATTS}(s, t) f(s) ds - \int_{s=0}^{\bar{\tau}-\tau'} \text{ATTS}(s, t) f(s) ds \quad (9)$$

Assuming that  $\tau'$  describes the more recent inflow cohort ( $\tau' > \tau$ ), then

$$\int_{s=0}^{\bar{\tau}-\tau} \text{ATTS}(s, t) f(s) ds - \int_{s=0}^{\bar{\tau}-\tau'} \text{ATTS}(s, t) f(s) ds = \int_{s=\bar{\tau}-\tau'}^{\bar{\tau}-\tau} \text{ATTS}(s, t) f(s) ds$$

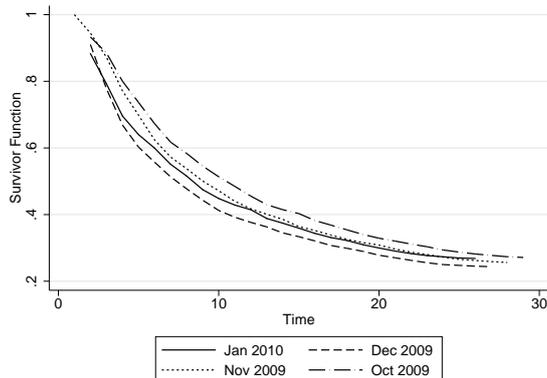
This simple cohort difference identifies the impact of program participation plus potential business cycle and seasonality differences. To eliminate the effect of seasonality we consider the inflow cohorts 12 months earlier, so  $\tau - 12$  and  $\tau' - 12$ . We take  $t$  such that both these inflow cohorts are up to  $t$  unaffected by the policy discontinuity, then

$$\begin{aligned} \left[ S_{\tau}(t) - S_{\tau'}(t) \right] - \left[ S_{\tau-12}(t) - S_{\tau'-12}(t) \right] &= b_{\tau}(t) - b_{\tau'}(t) - b_{\tau-12}(t) + b_{\tau'-12}(t) \\ &+ \int_{s=\bar{\tau}-\tau'}^{\bar{\tau}-\tau} \text{ATTS}(s, t) f(s) ds \quad (10) \end{aligned}$$

This double difference identifies the impact of program participation if we make the identifying assumption  $b_{\tau}(t) - b_{\tau-12}(t) = b_{\tau'}(t) - b_{\tau'-12}(t)$ . This assumption imposes that during the observation period the business cycle effects change only very smoothly. Recall that Figure 9 suggests that, if the observation period is sufficiently small, seasonal effects are larger than business cycle effects. Our double-difference estimator is an extension of the approach suggested by Van den Berg et al. (2014), who exploit a policy discontinuity to estimate effects on a duration variable. In Online Appendix C we discuss in more detail the assumption that business cycle effects are small (which is similar to the standard common trend assumption in difference-in-differences), and provide placebo estimates that support the assumption.

In our empirical approach we estimate the intention-to-treat effects, specified as  $\int_{s=\bar{\tau}-\tau'}^{\bar{\tau}-\tau} \text{ATTS}(s, t) f(s) ds$ , but using the empirical program participation rate  $f(s)$  it is straightforward to obtain  $\text{ATTS}(s, t)$ , which describes the average treatment effect on the treated survivors.

Figure 10: Survivor functions by month of inflow



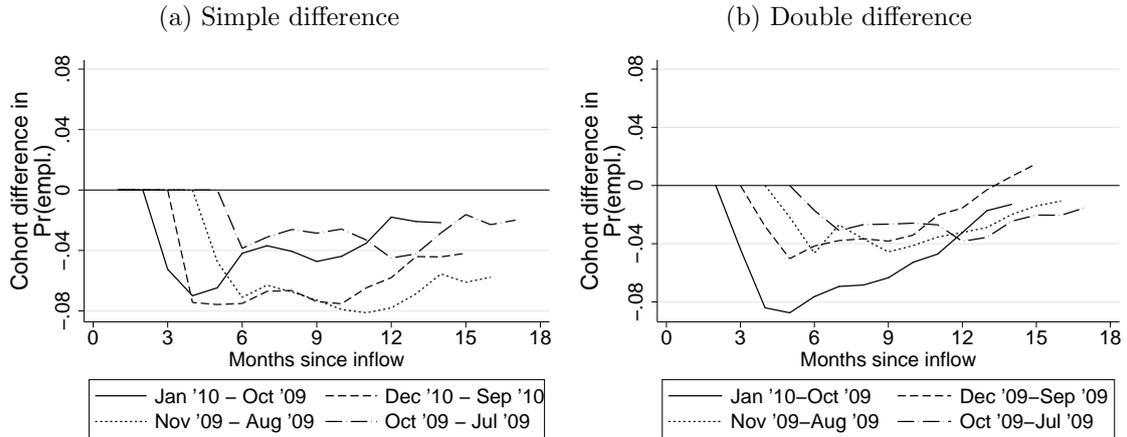
## 6.2 Estimation results

We start by defining which cohorts to compare. The time interval between cohorts should be small to minimize business cycle effects, but the trade-off is that more time between cohorts increases the difference in exposure to program participation. We use cohorts three months apart. To exploit the policy discontinuity, the cohorts should not enter unemployment too long before March 2010. Therefore, we use the cohorts of October 2009 until January 2010, facing between five and two months of potential program participation, respectively. Each cohort is compared to the cohort entering unemployment three months earlier. The survivor functions of each cohort are presented in Figure 10, showing that around 50% of the UI recipients find work within 12 months.

We first take the difference between the survivor function and the survivor function of the cohort entering unemployment three months earlier.<sup>23</sup> This compares the outflow in a cohort without enrollment in the program to the outflow in a cohort with regular enrollment in the program during the period before the policy discontinuity. We condition on survival and no-treatment up to the duration at which the later cohort reaches the policy discontinuity. The three-months differences are presented in panel (a) of Figure 11. We find very similar patterns across the different comparisons: a negative effect on job finding ranging from 3%-points to almost 8%-points which decreases in magnitude over time. The negative effects persist up to at least 13 to 17 months of unemployment.

<sup>23</sup>All estimates presented in this section are estimated using weights as discussed in Subsection 6.1.

Figure 11: Intention-to-treat effect estimates exploiting the policy discontinuity

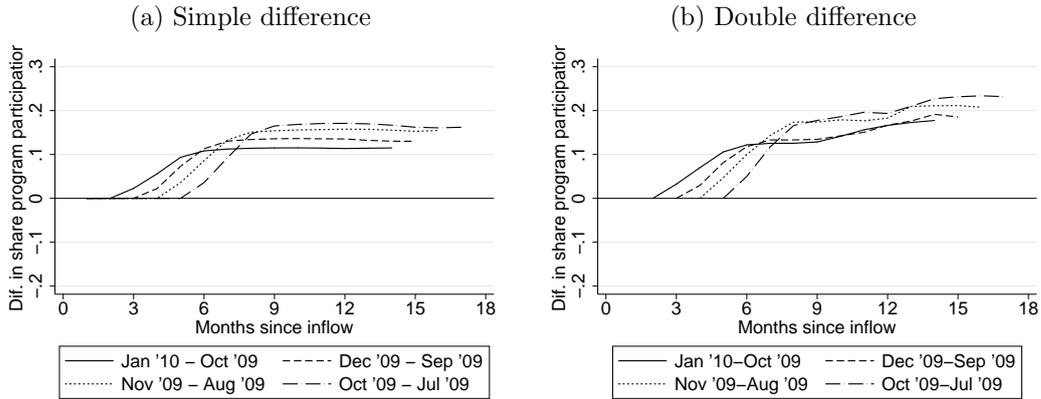


These estimates are based on simple differences between cohorts (equation (9)). By subtracting the same differences from a year earlier, we correct for seasonalities (equation (10)). Estimates from such a “difference-in-differences” approach are presented in panel (b) of Figure 11.<sup>24</sup> We find a very similar negative effect on job finding in the first months, although the reductions at longer durations are greater in magnitude than in the single difference estimates. At longer durations the estimates converge to a zero effect (or even a slightly positive impact). The effects are statistically significant for about the first 15 months (see Figures D12 and D13 in Online Appendix D for 95% confidence intervals). Each comparison measures the effect of treatment at a slightly different duration. For example, the January 2010-October 2009 comparison measures the effect of additional treatment in the third to fifth month of unemployment, while the December 2009-September 2009 comparison measures the effect of additional treatment in the fourth to sixth month of unemployment.

We find a negative impact of program participation on job finding that is consistent across different cohort comparisons and across the two estimators. This finding is in line with the lock-in effect that we have discussed earlier. These estimates measure intention-to-treat effects, and should be divided by the cohort differences in treatment participation to obtain average treatment effects.

<sup>24</sup>When estimating the effects we only present estimates up to the duration at which the cohorts from a year earlier reach the policy discontinuity, which is between 15 and 18 months. Estimates at longer durations are biased as the earlier cohorts are affected by the policy discontinuity as well.

Figure 12: Differences in program participation



We estimate the difference in program participation by

$$\int_{s=0}^{\bar{\tau}-\tau'} \hat{f}_{\tau'}(s)ds - \int_{s=0}^{\bar{\tau}-\tau} \hat{f}_{\tau}(s)ds \quad (11)$$

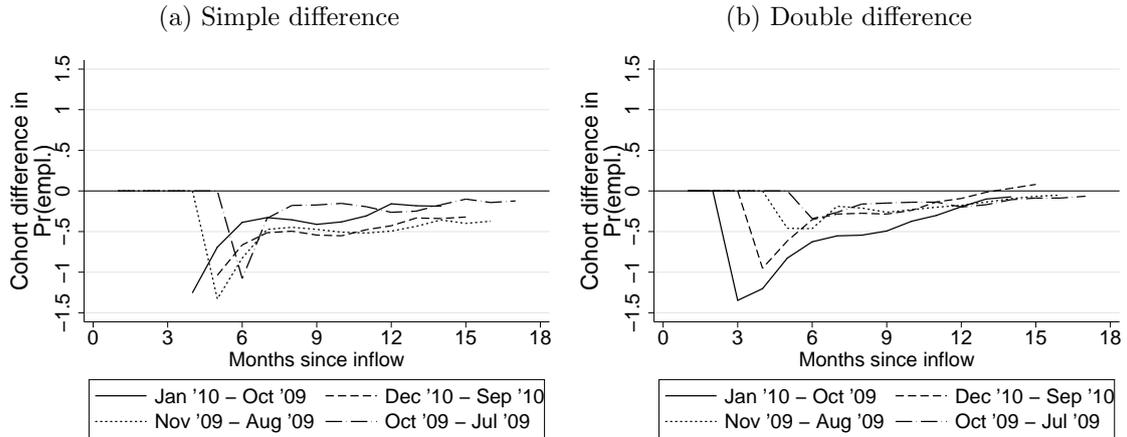
where  $\bar{\tau}$  is the moment of the policy discontinuity (March 2010) and  $\tau'$  and  $\tau$  are two inflow cohorts, e.g. January 2010 and October 2009. The functions  $\hat{f}$  describe the empirical program participation rates in these inflow cohorts. Panel (a) of Figure 12 shows estimated differences in program participation rates. There is a clear increase at the moment the first cohort reaches March 2010. The difference increases for approximately three months to about 10% to 20%-points. After that the comparison cohort reaches March 2010 and both cohorts no longer enroll in the program. The difference in program participation can also be computed using the same “difference-in-differences” approach, which are presented in panel (b) of Figure 12. The differences are less smooth, but exhibit the same pattern.<sup>25</sup>

The estimated ATET( $s, t$ ) are presented in Figure 13. Since differences in treatment participation between cohorts accrue over a three-months period, the initially small differences inflate the intention-to-treat effects substantially. To facilitate visual presentation, we remove some of the extreme estimates in the early months.<sup>26</sup> We find that the impact on the probability on being employed is around 20 to 50%-points, both when using simple differences (panel (a)) and double differencing (panel (b)). The effects decline over time. While a small negative impact remains in the

<sup>25</sup>All differences are highly statistically significant (see Figures D14 and D15 in Online Appendix D for confidence intervals).

<sup>26</sup>Complete graphs, including confidence intervals, are provided in Online Appendix D (see Figures D16) and D17).

Figure 13: Average treatment effect exploiting the policy discontinuity



simple-difference estimates, the impact seems to disappear in the double-difference estimates after about 15 months.

Note that from an identification perspective it is a major advantage that participation was reduced to zero (rather than to some positive value), because that implies that our estimates can be interpreted as average treatment effects on the treated, rather than local average treatment effects.<sup>27</sup>

## 7 Discussion

In this section we compare the estimated effects of the activation programs obtained using the three approaches.

Table 4 shows the estimated program effects of enrolling after  $s = 3, 4, 5, 6$  months on being employed after 12 months.<sup>28</sup> The pattern of program effects are very similar for the three methods, but the magnitudes of the estimates differ. Immediately after enrolling in the program, the job finding rate declines, and after some months the negative effect becomes smaller. The program thus postpones job finding, but does not affect the probability to have work in the long run.<sup>29</sup> This

<sup>27</sup>In the terminology of instrumental variables, there is full compliance with the instrument (the discontinuity) and there are no “always-takers”.

<sup>28</sup>The dynamic matching estimates coincide with those depicted in Figure 6, while the timing-of-events estimates coincide with Figure 7. For the quasi-experimental estimates, we include the double-difference estimates presented in panel (b) of Figure 13.

<sup>29</sup>We cannot predict what happens in the very long run beyond our observation period, but job search assistance programs (which constitute a large share of the activation programs that we consider) typically do not have very strong long-run effects (e.g. Card et al. (2010)).

Table 4: Comparison of estimation results across approaches

	(s=3, t=12)		(s=4, t=12)		(s=5, t=12)		(s=6, t=12)	
	coef	se	coef	se	coef	se	coef	se
Dynamic matching								
Full sample	-0.083	0.009	-0.089	0.009	-0.079	0.005	-0.068	0.009
Pre-discontinuity	-0.048	0.024	-0.110	0.011	-0.075	0.019	-0.049	0.017
Discontinuity	-0.064	0.028	-0.103	0.017	-0.114	0.021		
Timing-of-events								
Full sample	0.007	0.004	0.002	0.004	-0.004	0.003	-0.011	0.003
Pre-discontinuity	-0.010	0.005	-0.013	0.005	-0.017	0.004	-0.021	0.004
Discontinuity	-0.005	0.014	-0.010	0.013	-0.016	0.012	-0.023	0.011
Quasi-experimental								
Discontinuity	-0.195	0.064	-0.093	0.071	-0.177	0.081	-0.198	0.078
95% Conf. interval for differences: <sup>a</sup>								
Q-exp – Matching	[-0.24 ; -0.01]**		[-0.12 ; 0.20]		[-0.21 ; 0.11]			
Q-exp – T.-of-Events	[-0.31 ; -0.07]**		[-0.24 ; 0.08]		[-0.31 ; 0.03]**		[-0.31 ; 0.04]**	
Matching – T.-of-Events	[-0.12 ; -0.00]**		[-0.13 ; -0.05]**		[-0.15 ; -0.05]**			

All coefficients are estimates of  $ATTS(s, t)$ . <sup>a</sup>Confidence intervals (95%) of the difference between estimators (only for the discontinuity sample), computed using bootstrapping (100 repetitions). Stars indicate significance levels (\*\* 0.05, \* 0.10)

finding is also robust against using different samples and consistent with the results of Vikström (2017) for a Swedish work practice program. The common explanation for this pattern is the presence of lock-in effects (Lechner et al. (2011)), which cause that effects on employment are negative and decline immediately after starting a program. In the period that the program effects start increasing, the effect on job finding is dominant. The program effects decline for about two to three months, which can be considered as the lock-in period. Lechner et al. (2011) report that lock-in periods can be quite long and even longer than the lock-in period that we find.

In their meta-analysis, Card et al. (2011) classify evaluation study results as a positive, negative or zero impact. However, if one goes beyond the sign of the impact, we find substantial differences in the magnitudes of the estimated treatment effects. The quasi-experimental estimate is largest in magnitude. This approach estimates the average treatment effect by dividing the intention-to-treat effect by a small fraction of treated individuals. This causes that the treatment effect is estimated relatively imprecise. In the lower panel of Table 4 we present confidence intervals

for the differences between estimates from the different approaches (only for the discontinuity sample). These confidence intervals are computed using bootstrapping. We find that the quasi-experimental estimate for  $(s = 3, t = 12)$  differs significantly from the matching estimate, while this is not the case for  $s = 4$  and  $s = 5$ . The timing-of-events estimates are substantially smaller in magnitude than the quasi-experimental and matching estimates, and we find that these differences are mostly statistically significant.

The difference between the magnitudes of the estimates may suggest that not all identifying assumptions hold. First, the timing-of-events study finds some relevant unobserved heterogeneity. This may either imply that the conditional independence assumption required for the matching estimator might be violated or that the mixed proportional hazard specification in the timing-of-events model is too restrictive. Second, the common-trend assumption in the quasi-experimental approach could be violated, leading to a (small) downward bias. Finally, the no-anticipation assumption is required to define treatment effects. Due to the unexpected policy discontinuity this assumption is less likely to be violated in the quasi-experimental evaluation.

Our empirical analysis provides estimates for the  $ATTS(t, s)$ , which is the impact on individual employment probabilities for specific values of  $s$  and  $t$ . A policy maker might be interested in the average effect for all participants. We have defined the  $ATET(t)$  in equation (2), which is essentially an average impact of program participation weighted by the inflow in the program at duration  $s$ . Since all three identification strategies suggest a negative impact of program participation for any  $s$ , the corresponding  $ATET(t)$  will also be negative for all values of  $t$ . As illustration, we present the estimated  $ATET(t)$  for the dynamic matching estimates in Figure D18 in Online Appendix D. For any unemployment duration  $t$  participation in the program reduces the probability to have found work.

## 8 Conclusion

Several methods are available when evaluating activation programs for unemployed job seekers. In this paper we compare estimates from three different methods. First, we apply a dynamic matching estimator, second we estimate the timing-of-events model, and third we exploit exogenous variation in program participation, caused

by budgetary problems of the UI administration.

All three methods suggest a significantly negative effect of program participation on unemployment durations. The quasi-experimental estimates suggest reductions in the probability of being employed of up to 50%-points shortly after inflow in the program, the matching and timing-of-events estimates are smaller in magnitude (2.5 - 15%-points). In the longer run, all three methods suggest an (imprecise) zero effect on being employment. The robust conclusion drawn from each approach is that the programs are not effective in increasing outflow, and even reduce employment rates in the short and medium long run. We find some unobserved heterogeneity in the timing-of-events model which might explain the difference in magnitude compared to the dynamic matching estimates. However, it does not change the pattern of treatment effects between these methods. Our results concur with the meta-analysis performed by Kluge (2010), who does not find a relation between the methodology and the likelihood of estimating positive or negative effects. Our results confirm that this also holds when evaluating the same program using the same data, rather than when comparing across studies.

The policy discontinuity was caused by budgetary problems at the UI administration and the refusal of the Ministry of Social Affairs to extend the budget. This resulted in the termination of using privately provided activation programs, which can be considered as a policy measure to deal with the consequences of the ongoing economic crisis early 2010. Our finding that participation in privately provided activation programs does not shorten UI benefits periods, suggests that it was quite a successful policy measure to cut government expenditures. Our results also concur with other studies showing the lack of effectiveness of activation programs offered by commercial providers (e.g. Behaghel et al. (2014), Cottier et al. (2015) and Krug and Stephan (2016)). We have focused on employment as the key outcome variable. We cannot assess whether the program affects job quality (such as the salary or job stability), and thus we cannot exclude that perhaps it has a positive impact along these dimensions. However, using data on UI benefits recipients in the Netherlands, De Groot and Van der Klaauw (2019) and Van der Klaauw and Ziegler (2019) find that similar policy interventions have stronger effects on job finding than on wages and that effects on wages are mostly negative.<sup>30</sup>

---

<sup>30</sup>A back of the envelop calculation using our matching estimates suggests that program partic-

## References

- Abbring, J. H. and Heckman, J. J. (2007). Econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B, pages 5145–5303. Elsevier.
- Abbring, J. H. and Van den Berg, G. J. (2003). The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.
- Abbring, J. H., Van den Berg, G. J., and Van Ours, J. C. (2005). The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *Economic Journal*, 115(505):602–630.
- Behaghel, L., Crépon, B., and Gurgand, M. (2014). Private and public provision of counseling to job seekers: Evidence from a large controlled experiment. *American Economic Journal: Applied Economics*, 6(4):142–74.
- Bennemarker, H., Grönqvist, E., and Öckert, B. (2013). Effects of outsourcing employment services: Evidence from a randomized experiment. *Journal of Public Economics*, 98:68–84.
- Biewen, M., Fitzenberger, B., Osikominu, A., and Paul, M. (2014). The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897.
- Bléhaut, M. and Rathelot, R. (2014). Expérimentation contrôlée contre appariement: le cas d’un dispositif d’accompagnement de jeunes diplômés demandeurs d’emploi. *Economie & Prévision*, 204-205(1-2):163–181.
- Brodaty, T., Crépon, B., and Fougère, D. (2001). Using matching estimators to evaluate alternative youth employment programs: Evidence from france, 1986–
- 
- ipation extends the UI benefits period by about 3 weeks. De Groot and Van der Klaauw (2019) report that the median UI benefits period is about 100 days and wages in the first job are, on average, only 7% higher than the benefits level (Van den Berg and Van der Klaauw (2019) report a difference of about 9%). So if program participation increases wage by 1%, then the wage increase compensates workers within half a year for the extended benefits period.

1988. In Lechner, M. and Pfeiffer, F., editors, *Econometric Evaluation of Labour Market Policies*, pages 85–123, Heidelberg. Physica-Verlag HD.
- Card, D., Ibararán, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The labor market impacts of youth training in the Dominican Republic. *Journal of Labor Economics*, 29(2):267–300.
- Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*, 120(548):F452–F477.
- Card, D. and Sullivan, D. G. (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica*, 56(3):497–530.
- Cockx, B. and Dejemeppe, M. (2012). Monitoring job search effort: An evaluation based on a regression discontinuity design. *Labour Economics*, 19(5):729–737.
- Cottier, L., Kempeneers, P., Flückiger, Y., and Lalive, R. (2015). Does outsourcing job search assistance help job seekers find and keep jobs? IZA Discussion Paper.
- De Groot, N. and Van der Klaauw, B. (2019). The effects of reducing the entitlement period to unemployment insurance benefits. *Labour Economics*, 57:195–208.
- Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062.
- Dolton, P. and O’Neill, D. (2002). The long-run effects of unemployment monitoring and work-search programs: Experimental evidence from the United Kingdom. *Journal of Labor Economics*, 20(2):381–403.
- Fredriksson, P. and Johansson, P. (2008). Dynamic treatment assignment: the consequences for evaluations using observational data. *Journal of Business and Economic and Statistics*, 26(4):435–445.
- Gerfin, M. and Lechner, M. (2002). A microeconomic evaluation of the active labour market policy in Switzerland. *Economic Journal*, 112(482):854–893.

- Graversen, B. K. and Van Ours, J. C. (2008). How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program. *Journal of Public Economics*, 92(10-11):2020–2035.
- Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4):605–654.
- Heckman, J. J., Lalonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. C. and Card, D., editors, *Handbook of Labor Economics*, volume 3A, chapter 31, pages 1865–2097. Elsevier.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Kastoryano, S. and Van der Klaauw, B. (2011). Dynamic evaluation of job search assistance. Discussion Paper 5424, IZA Bonn.
- Kluge, J. (2010). The effectiveness of European active labor market programs. *Labour Economics*, 17(6):904–918.
- Krug, G. and Stephan, G. (2016). Private and public placement services for hard-to-place unemployed - an evaluation of a randomized field experiment using register data. *Industrial and Labor Relations Review*, 69(2):471–500.
- Lalive, R., van Ours, J. C., and Zweimüller, J. (2005). The effect of benefit sanctions on the duration of unemployment. *Journal of the European Economic Association*, 3(6):1386–1417.
- Lalive, R., Van Ours, J. C., and Zweimüller, J. (2008). The impact of active labour market programmes on the duration of unemployment in Switzerland. *Economic Journal*, 118(525):235–257.
- LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76(4):604–620.

- Lechner, M., Miquel, R., and Wunsch, C. (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 Wunsch, C. (2013). Sensitivity of matching-based program evaluations to the ability to control variables. *Labour Economics*, 21:111–121.
- Lombardi, S., Van den Berg, G., and Vikström, J. (2018). Empirical Monte Carlo evidence on estimation of timing-of-events models. Technical report, mimeo.
- Mueser, P. R., Troske, K. R., and Gorislavsky, A. (2007). Using state administrative data to measure program performance. *Review of Economics and Statistics*, 89(4):761–783.
- Muller, P., van der Klaauw, B., and Heyma, A. (2017). Comparing econometric methods to empirically evaluate job-search assistance. IZA Discussion Papers 10531, Bonn.
- Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86(1):133–155.
- Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde’s critique of nonexperimental estimators? *Journal of econometrics*, 125(1-2):305–353.
- Van den Berg, G. and Van der Klaauw, B. (2001). Combining micro and macro unemployment duration data. *Journal of Econometrics*, 102(2):271–309.
- Van den Berg, G. J., Bozio, A., and Dias, M. C. (2014). Policy discontinuity and duration outcomes. Discussion Paper 8450, IZA Bonn.
- Van den Berg, G. J. and Van der Klaauw, B. (2006). Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment. *International Economic Review*, 47(3):895–936.
- Van den Berg, G. J. and Van der Klaauw, B. (2019). Structural empirical evaluation of job search monitoring. *International Economic Review*, 60(2):879–903.
- Van den Berg, G. J., Van der Klaauw, B., and Van Ours, J. C. (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics*, 22(1):211–241.

- Van der Klaauw, B. and Van Ours, J. C. (2013). Carrot and stick: How re-employment bonuses and benefit sanctions affect exit rates from welfare. *Journal of Applied Econometrics*, 28(2):275–296.
- Van der Klaauw, B. and Ziegler, L. (2019). A field experiment on labor market speeddates for unemployed workers. Discussion Paper 12140, IZA.
- Vikström, J. (2017). Dynamic treatment assignment and evaluation of active labor market policies. *Labour Economics*, 49:42–54.