Evaluating the impact of a mandatory job search program: evidence from a large longitudinal dataset*

Luís Centeno
CEETA, Lisbon
lcenteno@telepac.pt

Mário Centeno
Banco de Portugal and ISEG
mcenteno@bportugal.pt

Álvaro A. Novo
Banco de Portugal and ISEGI
anovo@bportugal.pt

13th September 2004
Draft version

Abstract

We assess the impact on unemployment duration and re-employment wages of a large scale employment program implemented in Portugal during the late 1990's and early 2000's. Based on a dataset covering over 1.5 million individuals, we construct a pooled set of treatment and control groups to infer the effect of the program. We test the program effectiveness on two outcome variables: the unemployment duration and wages after unemployment. Based on the class of difference-in-differences matching methods, the estimates of the average effects of the treatment on the treated point towards: (i) a reduction in unemployment spells smaller than 1 month, rather small in the context of long unemployment spells observed in Portugal and (ii) typically a reduction, if statistically insignificant, of the re-employment wages.

Keywords: job search assistance; unemployment duration; policy evaluation
JEL Classification: J18, J23, J64, J68

*We thank participants at the 2004 LoWER Conference, London. The opinions expressed herein are not necessarily those of Banco de Portugal. All errors remain ours.

1 Introduction

It is a well-known fact in many European countries that unemployment duration is rather long, generating problems associated with long-term unemployment, such as weak labor market attachment. In Portugal, despite the low level of the unemployment rate, long-term unemployment is quite common, making the low unemployment level a social and economic issue. This pattern of unemployment duration can be seen as a trap, making long periods in that state increasingly harder to leave due to some sort of skills depreciation or because it sends a negative signal to the market about the unemployed.

In response to the high unemployment figures for specific labor market groups, such as young workers, women and those aged 45 or more, European Union countries increased their spending on active labor market policy and most countries implemented large scale programs targeted to these specific groups. In this paper, we assess the impact on unemployment duration of a large (mandatory) job search program implemented in Portugal during the late 1990's and early 2000's. The Portuguese program's goal was to increase the employability of the long-term unemployed (the REAGE program) and to act earlier on youth unemployment preventing episodes of long-term unemployment at the beginning of their labor market career (the InserJovem program).

The objective of this paper is to determine the effects of the programs comparatively to the outcome of the individual had he/she continued to search for a job in the absence of the job search support provided by the program. The focus is on the direct effects of the programs; no attempt is made to assess the general equilibrium implications. However, we should stress that displacement effects, the ones that we would expect to identified in general equilibrium approaches, are more likely in employment assistance program (for example, with wage subsidies) than through the kind of job search programs we are evaluating here.

With this objective in mind and the estimation issues that arise in non-experimental studies, mainly due to the problem of missing data, we select a set of methods developed to address such settings. The methodologies used include: (i) matching methods (Heckman, Ichimura & Todd 1998), in which the propensity score matching is based on different definitions of neighbor; (ii) difference-in-difference-in-differences (Meyer 1995) modeling strategy to tackle the problems associated with selection on non-observables and (iii) and as proposed by Heckman, Ichimura & Todd (1997) and Heckman, Ichimura, Smith & Todd (1998) a combination of the two methodologies, termed difference-in-differences matching, is also used to eliminate potential sources of bias (Smith & Todd 2004). In the same spirit, we extend Meyer (1995) approach by combining it with the matching estimators.

Previous microeconometric studies of active labor market programs include Blundell, Dias, Meghir & Reenen (2004) and Larsson (2003). The results of these studies are mixed. Whereas Blundell et al for the UK find an important "program introduction effect", the program effect is much larger in the first quarter than later on, Larsson finds no significant effect in the Swedish programs, and if any thing she finds a negative impact of certain aspects of the program.

Our results point out to a small, non-significant, effect of the programs on unemployment duration and sometimes negative impact on wages. We estimate that unemployment duration is reduced by at most 1 month, which does not represent a large decrease in duration given that some workers can spend several years unemployed. In terms of wages, after leaving unemployment, we estimate a null impact for women and a large negative impact on males' re-employment wages.

The paper is organized as follows. The evaluation problem, as well as the identification and estimation of average treatment effects is addressed in section 2. The labor market program is described in Section 3. Section 4 presents the data and results. Finally, concluding remarks are presented in Section 5.

2 Econometric evaluation strategies

The problem of evaluating active labor market programs has been extensively studied in the literature (Heckman, LaLonde & Smith 1999). In recent years, a wealth of methods to address the main problem of missing data present in all non-experimental studies has been proposed. These methods suggest different solutions to the problem of generating conveniently designed comparison groups necessary to perform program evaluations. Given the non-experimental feature of these programs, the feasibility of any evaluation exercise depends crucially on the ability that researchers have to generate such comparison groups from the data available on the program implementation. The methodologies proposed include: (i) matching methods (Heckman, Ichimura & Todd 1998), in which the propensity score matching is based on different definitions of neighbor; (ii) regression methods – in particular, duration models to estimate the shift in the unemployment hazard attributable to the program, and, finally, (iii) difference-in-difference-in-differences (Meyer 1995) modeling strategy to tackle the problems associated with selection on non-observables. More recently, Smith & Todd (2004) proposed a combination of the propensity score matching with the difference-in-differences, resulting in the difference-in-differences matching estimator. In their study, this estimator appeared to eliminate potential sources of bias.

2.1 The evaluation problem

Active job search programs have the objective of easying/speeding the transition from unemployment to employment. The evaluation problem faced in this study is precisely to measure the impact of such a program in the duration of unemployment spells of the target population – young experiencing short-lived unemployment (under 25 years and unemployment spell above 6 months) and older individuals with long-term unemployment (over 25 years old and spell over 12 months).

As with all non-experimental cases, this study also faces the problem of missing data in the evaluation problem. The fact that the same individual is not observed simultaneously in the two states – treatment and control – conditions the entire evaluation. It is, therefore, necessary to devise sampling strategies and use methods that attempt to overcome such observational limitations, which may have deeper statistical consequences (e.g. due to self-selection, Heckman (1979)) in the estimation process.

2.2 Econometric Methods

2.2.1 Matching as an evaluation estimator

One of the most important issues in the evaluation of non-experimental employment programs is that of the comparision of individuals who are indeed comparable (Rubin 1977). How such concern is addressed is key to identifying the effect of the treatment.

For that purpose, the selection of the control group can be achieved with recourse to the non-parametric method of matching in the *propensity score*. In short, the methodology compresses the indivuals' observable multivariate information, and as such difficult to compare, in an indicator – the propensity score. The distance between the individual propensity scores is minimized to construct a control group comparable in the observables with the treatment group (Rubin 1977, Rosenbaum & Rubin 1983).

The propensity score is defined as the conditional probability, on the observable pre-treatment characteristics, of receiving treatment. Formally,

$$p(X) = \Pr[T = 1|X] = E[T|X], \tag{1}$$

where T is a binary variable equal to 1 if treatment is received and 0 otherwise and X is a vector of observable characteristics. The selection of the control in $p(X)$, rather than X, is possible if the following conditions are verified:

1. *Balancing of score.*

$$T \perp X \mid p(X). \tag{2}$$

2. *Unconfoundness.* If

$$Y_1, Y_0 \perp T \mid X, \tag{3}$$

then

$$Y_1, Y_0 \perp T \mid p(X). \tag{4}$$

Intuitively, the first condition requires that given the propensity score, exposition to the treatment (or not) is independent of the observable characteristics (X). Relatively to the unconfoundness condition, this states that if, conditional on X, the treatment status is independent of the expected outcome resulting from treatment Y_1 and from no treatment Y_0, then conditioning on $p(X)$ preserves this important condition of independence. While the first condition can be tested, and the samples redefined until it is verified, the second condition cannot be tested.

$p(X)$ is computed with regression models for binary variables, such as the probit and logit. Because propensity scores are continuous variables, the probability that any two $p(X)$ are equal is zero, requiring the use of matching methods to construct comparable treatment and control groups. The most commom matching methods are (i) stratification matching, (ii) nearest neighbor matching, (iii) radius matching and (iv) kernel matching.[1] Applications of these statistical methods to the evaluation of active labor market programs can be found in Blundell & Dias (2000) and Larsson (2003).

2.2.2 Difference-in-difference-in-differences

The above methods control for observed characteristics to infer the impact of the treatment. However, there are situations in which the treatment and control groups differ in terms of unobserved characteristics, resulting in two groups that are not fully comparable. In such circunstances, the difference-in-differences method offers a possible solution.

Let Y_{it}^D be the potential outcome for individual i at time t given that he/she is in state D, where $D = 1$, if the individual received treatment and 0 otherwise. Let treatment take place at time $t = 1$. The fundamental identification problem lies in the fact that we do not observe, at time $t = 1$, individual i in both states. Therefore, we cannot compute the individual treatment effect, $Y_{i1}^1 - Y_{i1}^0$. One can, however, estimate the average effect of the treatment on the treated, $E[Y_{i1}^1 - Y_{i1}^0 \mid D = 1]$. In order to achieve

[1] See Becker & Ichino (2002) for a concise review of each of propensity score matching methods and an estimation algorithm.

identification, the following assumption is fundamental:

$$E[Y_{i1}^0 - Y_{i0}^0 \mid D = 1] = E[Y_{i1}^0 - Y_{i0}^0 \mid D = 0] \tag{5}$$

It states that the temporal evolution of the outcome variable of treated individuals ($D = 1$) in the event that they had not been exposed to the treatment would have been the same as the observed for the individuals not exposed to the treatment $D = 0$. If the assumption expressed in (5) holds, then the average treatment effect on the treated can be estimated by the sample analogs of

$$\{E[Y_{i1} \mid D = 1] - E[Y_{i1} \mid D = 0]\} - \{E[Y_{i0} \mid D = 1] - E[Y_{i0} \mid D = 0]\}. \tag{6}$$

There are two threats to the validity of the difference-in-differences estimator. First, if cross-sectional data are used, compositional changes over time may invalidate the results. Second, if the dynamic evolution of the outcome variable depends on non-observables, identification breaks down.

Meyer (1995) proposed a difference-in-difference-in-differences approach that attempts to further correct for non-observables that threatment the validity of the difference-in-differences estimator. Suppose that the data have two dimensions – space and time. Then, it is possible to control for a space-time dimension. For instance, consider Table 1 with a time dimension (before-after) and a space dimension (treatment region and no-treatment region), where Y_{Rt}^D stands for the outcome of group D, treatment or control, in space R, treatment or no-treatment regions, at time t.

Consider the top panel of Table 1. Reading in row, the difference between the two rows gives a first measure of the impact of the treatment on the treated. That is, it corrects the outcome variable evolution of treated individuals (1st row) with the effect on pseudo-treated (same eligibility criteria) observed in a different space $R = 0$; it corrects for common factors influencing the target group.[2] The D-in-D estimator is $Y_{1t'}^1$-Y_{1t}^1-$Y_{0t'}^1$+Y_{0t}^1.

It is, however, possible that there are effects/shocks specific to the space of implementation, which although related with the treatment are not directly attributable to the program. The bottom panel of Table 1 estimates such effects by performing the same exercise as in the top panel, D-in-D, but now for a control group (different eligibility criteria). This measure, $Y_{1t'}^0$-Y_{1t}^0-$Y_{0t'}^0$+Y_{0t}^0, is a measure of the non-observables affecting the D-in-D estimator computed in the region of implementation of the treatment. Thus, the corrected effect of the treatment is given by DDD estimator $Y_{1t'}^1$-$Y_{0t'}^1$-Y_{1t}^1+$Y_{0t}^1 - Y_{1t'}^0$+$Y_{0t'}^0$+Y_{1t}^0-Y_{0t}^0.

[2]This resembles the estimation strategy of the matching estimators.

Table 1: Difference-in-difference-in-differences estimator (DDD)

Group	Region	Before (t')	After (t)	
Treatment	R=1	$Y^1_{1t'}$	Y^1_{1t}	Y^1_{1t}-$Y^1_{1t'}$
	R=0	$Y^1_{0t'}$	Y^1_{0t}	Y^1_{0t}-$Y^1_{0t'}$
		$Y^1_{1t'}$-$Y^1_{0t'}$	Y^1_{1t}-Y^1_{0t}	Y^1_{1t}-$Y^1_{1t'}$-Y^1_{0t}+$Y^1_{0t'}$
Control	R=1	$Y^0_{1t'}$	Y^0_{1t}	Y^0_{1t}-$Y^0_{1t'}$
	R=0	$Y^0_{0t'}$	Y^0_{0t}	Y^0_{0t}-$Y^0_{0t'}$
		$Y^0_{1t'}$-$Y^0_{0t'}$	Y^0_{1t}-Y^0_{0t}	Y^0_{1t}-$Y^0_{1t'}$-Y^0_{0t}+$Y^0_{0t'}$
		$Y^1_{1t'}$-$Y^1_{0t'}$-$Y^0_{1t'}$+$Y^0_{0t'}$	Y^1_{1t}-Y^1_{0t}-Y^0_{1t}+Y^0_{0t}	
DDD Estimator				$Y^1_{1t'}$-$Y^1_{0t'}$-Y^1_{1t}+Y^1_{0t} -$Y^0_{1t'}$+$Y^0_{0t'}$+Y^0_{1t}-Y^0_{0t}

Note: Superscripts denote treatment status; subscripts refer to space and time.

2.2.3 Difference-in-differences matching

Heckman et al. (1997) and Heckman, Ichimura, Smith & Todd (1998) introduced the D-in-D matching estimator. This estimator combines the two previous approaches, which according to a recent study of Smith & Todd (2004) has the potential to reduce sources of bias in non-experimental settings. Intuitively, the benefits may arise from the fact that: (i) relatively to the propensity score matching estimator, the D-in-D matching adds the control for non-observables that characterizes the D-in-D estimator, while (ii) relatively to the D-in-D estimator, its matching version adds the comparability on observable that characterizes the propensity score matching estimator.

Depending on the type of data available, the estimator can take two forms. With cross-sectional data, propensity score estimates of the average treatment effect on the treated for each period – before and after – are computed and then the difference between these two estimates yields the D-in-D matching estimates. If longitudinal data is available, the process is somewhat reversed. First, for each individual a difference between the after and before outcomes is computed and then using the propensity score matching each treatment unit is matched to control unit(s) yielding an average treatment effect on the treated.

In the empirical section, we will use this estimator, extending it to the DDD estimator. In particular, we will match the treatment units with the control units before computing the DDD, i.e., we will compute two D-in-D m matching estimates before taking the final difference.

3 The program: description

We study a large-scale program, implemented in Portugal in the context of the European Employment Strategy. Similar programs have been subject to evaluation in other countries (see, for example, Larsson (2003), for a study of the Swedish Youth Practice Program and Blundell et al. (2004) for the British New Deal Program). The Portuguese program is fundamentally a job search support program and its main goal is to improve the employability of two specific groups of unemployed individuals: those aged less than 25 years old and unemployed more than six months (the Programa InserJovem) and those over 25 and unemployed longer than 12 months (the Programa REAGE). The program was launched in a period of quite low unemployment rate, but with a poor performance of the unemployed in terms of duration, and specific demographic groups were doing poorly in the labor market, namely the less educated young and females. These two groups are more likely to be registered in the National Employment Offices, representing a large share of our sample.

The program is composed of intensive job-search assistance and small basic skills courses. Each individual is enrolled in a number of interviews with placement officers to help her improve her job-search skills. In case the placement team considers necessary, the individuals can enter a number of vocational or non-vocational training courses. The whole process of job-search assistance ended in most cases, but not necessarily, with the elaboration of a "Personal Job Plan", that included detailed information on the unemployed job search effort. According to this Plan, the unemployed was expected to meet on a regular basis with the placement officer at the local employment office and to actively search for a job. Unjustified rejection of job offers lead to a cancellation of any subsidies being received by the unemployed. The program is mandatory in the sense that failing to comply with it results in a cancellation of the workers registration and possibly of all subsidies he might have been receiving.

The design of the program does not exclude the possibility of spillover effects from its implementation to other workers not participating in the program. However, these general equilibrium effects are likely to be small given the nature of the program, for example, there are no wage subsidies.

4 Empirical results

We study the impact of the program on the average unemployment spell duration and on the average wages after re-employment. In both cases, we begin the analysis by using propensity scores to match treatment and control groups, controlling for observed characteristics. Then, we use two layers of control for unobserved characteristics. First, the difference-in-difference estimator and, finally, by taking advantage of the wealth of the database, introduce an additional layer of control by computing the difference-in-difference-in-differences estimator.

An explicit aim of active labor market policies is to improve the employability of the unemployed. Hence, a shorter unemployment duration, a higher probability of future employment or higher employment attachment – that can operate through better matches, and higher earnings – are possible measures of a program's success. We begin by analyzing the impact of the program on unemployment duration.

4.1 Unemployment duration

4.1.1 Description of the data

The Portuguese employment agency collected data from all registered unemployed regardless of their "treatment status". The dataset, comprising over 2 million observations for over 1.5 million individuals, monitors the different features of the program and individuals during their complete spells of unemployment. The information in the dataset includes most demographic variables used in labor market studies (age, sex, nationality, schooling, place of residence), including also a large number of variables related with previous labor market experience (previous occupation, desired sector of employment, unemployment duration, reason for job displacement). The unemployed is observed for the complete duration of the unemployment spell and, at the moment of termination, we can observe the destination state (either employment, training or out of the labor force). Additionally, we have some information of the vacancies posted in the labor office, which are possibly matched with each unemployed.

The main limitations of our data are the lack of wage information, which we try to construct using a different dataset in the next subsection, and the difficulty in following up the individuals after they leave the program. In fact, there is no income-related information at the Employment Services and follow up interviews were not carried out (although they were part of the original program design).

4.1.2 Sample construction

We take advantage of the wealth of information the dataset contains and of the characteristics of the program to construct treatment and control groups using different criteria. In particular, we explore (i) the longitudinal nature of the data, and (ii) the two sources of

Table 2: Summary statitics local Employment Offices

	Treatment group		Control group	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Age (in years)	31.9269	12.8478	33.3517	43.6103
Sex (Male =1)	0.3728	0.4835	0.4139	0.4941
UI recipient	0.2259	0.4182	0.2831	0.4505
Foreigner	0.0065	0.0805	0.0152	0.1223
Disabled	0.0073	0.0850	0.0075	0.0863
Marital status				
Married	0.4800	0.4996	0.4867	0.4998
Divorced	0.0087	0.0931	0.0101	0.1000
Other	0.0253	0.1569	0.0258	0.1586
Single	0.4747	0.4994	0.4658	0.4988
Widow	0.0114	0.1061	0.0116	0.1069
Schooling				
4 years	0.2780	0.4480	0.2785	0.4482
6 years	0.2408	0.4276	0.2249	0.4175
9 years	0.1708	0.3763	0.1718	0.3772
11 years	0.0926	0.2898	0.0995	0.2993
12 years	0.0978	0.2971	0.0977	0.2970
3 years college	0.0254	0.1572	0.0269	0.1619
5 years college	0.0277	0.1642	0.0449	0.2072
Master	0.0001	0.0075	0.0001	0.0116
Ph. D.	0.0000	0.0000	0.0000	0.0045
Illiterate	0.0296	0.1695	0.0214	0.1448
No school, reads	0.0373	0.1894	0.0341	0.1815
Reason to register				
Student	0.1139	0.3176	0.1027	0.3036
Finished school	0.0613	0.2398	0.0510	0.2200
Finished training	0.0095	0.0973	0.0045	0.0667
Worked at home	0.0136	0.1160	0.0148	0.1207
Laid-off	0.2004	0.4003	0.2624	0.4399
Quit job	0.0321	0.1762	0.0350	0.1838
End job mutual aggreement	0.0164	0.1270	0.0264	0.1605
End of temporary job		0.4701	0.2854	0.4516
Retired	0.0004	0.0212	0.0005	0.0217
Bad working conditions	0.0062	0.0787	0.0048	0.0690
Previoulsy self-employed	0.0001	0.0114	0.0001	0.0118
New registration	0.0021	0.0459	0.0028	0.0527
Law 17	0.0008	0.0277	0.0005	0.0231
Other	0.2133	0.4096	0.2090	0.4066
Number of observations				
By reason to exit				
Placed by program				
Cancelled registration				
Total	53407		201149	

variation coming from the eligibility criteria and the different implementation phases (which generate spatial/regional and time differences).

The program design and implementation generated a natural way to construct treatment and control groups along several dimensions. One such dimension is the eligibility criteria (based on age and unemployment duration) and the other is the phased implementation of the program across the Portuguese territory that generated a sequence of pilot areas. The program implementation was gradual. The local offices of the Employment Services were assigned to the program at different moments in time, divided by June and October of 1998, February, May, July and November of 1999 and by April, June and September of 2000.

The treatment group includes all individuals eligible to participate in the InserJovem and REAGE programs in the first six months of their implementation in each Employment Office. This generates a large group of individuals already unemployed at the moment the programs were initiated in each office (see Table 2 for the dimensions of treatment and control groups).

The choice of the comparison group was dictated by the eligibility criteria and the location outside the initial pilot areas. Thus the comparison group takes all eligible individuals living in the area covered by local Offices that did not implement the programs. This allows us to have nine possible comparisons between treatment and control groups that correspond to each of the nine dates at which different local offices started implementing the programs.

The timing of program implementation can be considered random. It was not dictated by special labor market conditions at the regional level (for example, regions of higher unemployment). This implies that the treatment and control groups, can be though of as a random draw from the Local Employment Offices at any point in time.

In Table 2, we present summary statistics for the two groups of interest. The two groups are not very different according to the characteristics presented in the table; however, there are some differences. Treated individuals are slightly younger, and they are more likely to be female. Among treated individuals the share of unemployment insurance (UI) recipients is smaller. The control group has a higher level of education, but the two groups are not very different along this dimension. The largest differences can be found in the "reason to register" attribute. The unemployed that were subject to treatment were more likely to have ended a temporary job than those in the control group, who are much more likely to have been laid-off prior to registration. These summary statistics are reassuring in terms of our ability to match individuals in the two groups to perform our evaluation exercise.

4.1.3 Unemployment duration impact

Our results, using the two econometric methods presented above, suggest a negligible impact in the employability of those receiving treatment (youth unemployed and older long term unemployed). The impact on the average unemployment spell ranges from a reduction of approximately 1 month to a slight increase of about 0.2 months. The analysis by gender and type of exit from registered unemployment reveals some differences, but still the impacts are rather minor. While younger males tend to benefit more than younger females, older females benefit the most from the treatment.

Difference-in-differences matching

The implementation of the matching method follows Becker & Ichino (2002), while the D-in-D matching estimator follows Smith & Todd (2004). First, using the fact that over time there are nine distinct events, we define at each event the treatment group of interest. Then, from the pool of individuals registered in local Employment Offices not yet in the programs, we construct the control group. The nine sets of treatment and control groups are then pooled to form the final larger treatment and control groups. Due to the heterogeneity of the individuals in each of the groups and, not independently, the fact that there are two programs – Inserjovem and Reage –, we splitted the sample into these two subsamples. The two subsamples are then analyzed according to: (i) the type of exit from the pool of registered unemployed – all exits, placed and canceled[3] and (ii) the gender – female, male and all. Therefore, in total there are 18 estimates

The propensity score matching results are based on the stratification method. The matching process typically led to balanced treatment and control groups.[4] For these two groups (matched on the observables, implying a common support) the mean unemployment duration effect (treated minus control average duration) is computed for a period before the program implementation and afterwards.[5] The differences between the after and before propensity score estimates of the average treatment effect on the treated yields the D-in-D estimates.

[3]The exit category *placed* includes all individuals who either through the program or by themselves are reported has having been placed in the labor market or in a training program; the exit category *canceled* includes all individuals who saw their registration canceled by the Offices due to having failed to fulfill one or more criteria.

[4]For the entire set of estimates presented below, there are some cases where the two groups are unbalanced. However, because the balancing property failed to hold mainly due to statistically, but not economically different averages of the age variable, we proceeded as if the two groups were balanced. The differences in average age were typically of a few months, which clearly do not affect the required economic comparability of two groups. These balancing property difficulties tended to arise more often in the Reage program analysis.

[5]The results are not reported in Table 3, but they are available from the authors upon request. Consequently, the number of observations used per estimate are not reported. However, the minimum number of observations on the treatment group is XXX and YYY in the control group.

Table 3: D-in-D matching and DDD matching estimates

			All		Male		Female	
		Exit	Effect	s.e.	Effect	s.e.	Effect	s.e.
Inserjovem	D-in-D for PEP=1	All	-0.15	0.096	-0.22	0.146	-0.11	0.126
		Placed	0.18	0.211	-0.04	0.300	0.21	0.278
		Canceled	-0.36	0.115	-0.38	0.175	-0.35	0.151
	D-in-D for PEP=0	All	0.00	0.013	-0.01	0.020	-0.01	0.017
		Placed	0.00	0.026	-0.03	0.039	0.02	0.033
		Canceled	-0.02	0.016	-0.02	0.025	-0.02	0.021
	DDD	All	-0.15	0.097	-0.21	0.147	-0.10	0.128
		Placed	0.18	0.213	-0.01	0.303	0.20	0.280
		Canceled	-0.34	0.116	-0.36	0.177	-0.33	0.152
Reage	D-in-D for PEP=1	All	-0.54	0.165	-0.48	0.222	-0.75	0.231
		Placed	0.09	0.330	0.38	0.428	-0.04	0.471
		Canceled	-0.56	0.197	-0.42	0.265	-0.89	0.278
	D-in-D for PEP=0	All	0.21	0.020	0.21	0.033	0.20	0.026
		Placed	0.29	0.035	0.31	0.057	0.29	0.044
		Canceled	0.10	0.027	0.05	0.045	0.13	0.034
	DDD	All	-0.75	0.166	-0.69	0.225	-0.95	0.232
		Placed	-0.20	0.332	0.06	0.431	-0.33	0.473
		Canceled	-0.66	0.199	-0.47	0.269	-1.01	0.280

In Table 3, the panels labeled 'PEP=1', indicating that the individuals participated or had the potential to participate in the programs, present the results. They show very minor positive effects of the two programs, which translate into marginally smaller unemployment durations for the treated. In Figure 1, the first column of plots summarizes the results of the difference-in-differences matching estimator, providing in addition to Table 3 the 95% confidence intervals. The following results are worth highlighting:

i) The impact of the program Inserjovem is lower than Reage's. Furthermore, the impact on the youth is statistically (and economically) insignificant. This result confirms previous findings in the literature (e.g. Blundell et al. (2004), Larsson (2003));

ii) The gender effect is stronger for females in Reage, that is, the D-in-D estimates are either more negative (or less positive), resulting in durations smaller than those observed for males. In the younger population, the differences are rather insignificant between the two sexes, but slightly in favor of men;

iii) In terms of the type of exit, the results are mixed, highlighting the importance of such disaggregation. Thus, when analyzing exits from the pool of registered unemployed due to either i) self-placement, ii) training or iii) placement through the local offices' efforts, which we aggregate in the class "placed", the D-in-D estimates for both programs are typically positive, if statistically insignificant. That is, the impact on duration is negligible, reaching in the best case a reduction of -0.04 months and in the worst case an increase of 0.38 months. When analyzing the group of individuals who exited the system due to failing to fulfill one or more criteria – class "canceled" – the estimates are negative and statistically significant. Somehow, the system seems to be more aware of the "irregularities", taking earlier action. Whether this is a desirable result is questionable – it may have a positive impact on the individuals affected, leading them to correct the behavior, or cause (further) stigma.[6] Overall, pooling all types of exits, the programs seem to have reduced unemployment duration, but only statistically significant for the Reage program, although the impact is of about only half month less.

Next, we control for other potential non-observable effects.

Selection on unobservables difference-in-difference-in-differences

The remaining panels of Figure 1 and Table 3 present the results from a difference-in-difference-in-differences (DDD) application. There are two sources of differences in this study. One is related with the eligibility to the program – based on age and unemployment duration. The other is related with the timing of implementation of the program, which at any point in time covers only a fraction of the offices of the national employment agency. Thus, first we consider only the local employment offices that joined the program at a particular moment in time and we construct treatment groups. Then, for the same time period, same eligibility criteria, but at non-eligible locations, we construct a pseudo-treatment group. Thus, we can construct two D-in-D matching estimates – one regarding the actually treated, which we discussed in the previous subsection and one regarding the pseudo-treated/control groups. The latter results and the difference between the two D-in-D matching estimates, the DDD matching estimates, are reported in the panels labeled 'PEP=0' and 'DDD' of of Figure 1 and Table 3.

The motivation to introduce an additional layer of control for unobserved effects is that there is the possibility that at the time of implementation of the program there were shock to the "treatment" areas that, even though might not be directly attributable to the program, are correlated with the program. For example, the simple awareness of the program might increase the firms' willingness to hire. Thus, we create another level of control. For that purpose, we repeat the D-in-D matching procedure, but this time we consider only the individuals that do not meet the eligibility criteria to participate in the program. For this group the older individuals actually saw their unemployment durations increase by 0.1 to

[6] This may be analyzed with a counting process method to assess the impact on the number of registrations per individual (exit/re-enter) in the system.

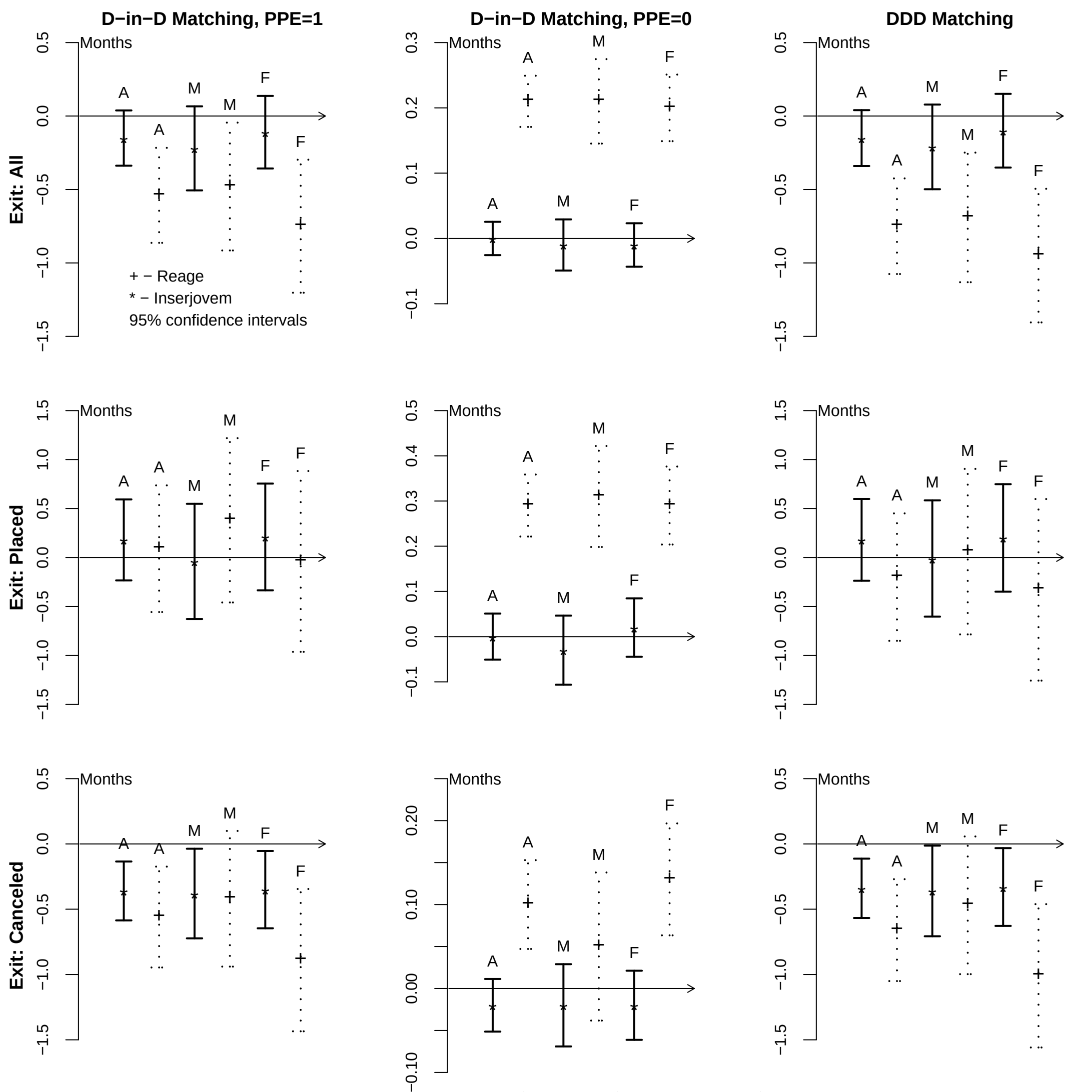

Figure 1: D-in-D matching and DDD matching

0.3 months. Younger individuals' durations before and after the program implementation are about the same, resulting in a null estimate of the average treatment effect.

Finally, the DDD matching estimates provide us with a corrected estimate for the average treatment effect on the treated, which is slightly better for the older individuals, but about the same for the younger population. The estimates and 95% confidence intervals are reported in the last column of Figure 1. Qualitatively, gender and type of exit effects described earlier hold true.

4.2 Wages: program's impact

In the previous analysis we used the reduction in the duration of unemployment as a measure of the program's success. As aforementionded, another obvious and important candidate is higher earnings. In our application earnings are measured as a continuous variable in monthly and hourly terms, using an alternative dataset.

4.2.1 Data source

We collect the earnings data from a different dataset, namely the Inquérito ao Emprego (IE), the Portuguese quarterly labor force survey. This is a unique dataset that samples the entire Portuguese population. The survey sample has more than 40,000 individuals each quarter and very detailed information on demographics and labor market status. As is common with other European labor force surveys individuals are interviewed in six consecutive quarters allowing us to create a short panel and observing transitions out of unemployment during the observation period.

Of special interest for our application is the information on unemployed workers searching behavior. Among the information available for unemployed workers we know whether they are registered at the Public Employment Services, a crucial variable to identify the treatment and control groups as we did with the administrative data. The implementation of the job search program described in section 2 guarantees that registered unemployed satisfying the two eligibility criteria (age and unemployment duration) were automatically enrolled in the program. This feature of the program allows us to identify enrolled (i.e. treated) unemployed in Employment Offices applying the program and the correct control group, either in Employment Offices not applying the program or individuals that would be enrolled before the program implementation, if the same program would have been in place.

We use the available sub sample of registered unemployed of the IE to study the program impact on wages. For unemployed and registered individuals, we are able to observe transitions out to employment because individuals are surveyed for six successive quarters.

Table 4: Summary statitics: Inquérito ao Emprego sample

Variable	Registered in t		Employed in $t+1$	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	38.96	13.84	33.18	11.80
Sex (1=Male)	0.38	0.49	0.40	0.49
Region = Lisbon	0.26	0.44	0.26	0.44
Region = North	0.35	0.48	0.32	0.47
Region = Center	0.10	0.30	0.08	0.27
Region = Alentejo	0.19	0.39	0.18	0.38
Region = Algarve	0.10	0.30	0.16	0.37
Education = No formal schooling	0.10	0.30	0.06	0.24
Education = $\leq$ 9 years	0.75	0.43	0.73	0.45
Education = 12 years	0.10	0.30	0.11	0.31
Education = Technical studies	0.02	0.14	0.03	0.17
Education = College degree	0.04	0.20	0.08	0.27
Unemployment insurance recipiency	0.36	0.48	0.33	0.47
Previous work experience (1=Yes)	0.88	0.32	0.88	0.32
Log Wage	–	–	11.23	0.33
Hours worked	–	–	40.01	7.38
Number of observations	18382		1194	

For those moving into employment we know their wage and hours of work. The survey has very detailed and accurate information on both variables, with monthly net wages recorded in nominal terms.

This information makes it possible to reproduce now for wages the matching on the propensity score difference-in-differences analysis carried out in the previous section. We obtain the treatment effect on starting wages of treated individuals.

4.2.2 Sample characterization

Before going into details with the results of the average treatment effect on wages we will briefly comment on some descriptive statistics from the IE sample, Table 4. The random sample obtained from IE is not much different from the one we obtained directly from the administrative data. Registered unemployed are on average young individuals, but the subsample of re-employed workers is on average five years younger. As happened with the SIGAE data, registered unemployed are mainly women and low educated workers. Most registered unemployed do not receive unemployment insurance, although the vast majority had previous work experience.

4.2.3 Results of the application to wages

Aggregating the pairwise differences over the common support yields an estimate of the average treatment effect on the treated. We are estimating differences in matching effects and Table 5 reports the estimated average treatment effect on the treated for the outcome variable wages. The left panel (PEP=1) presents results for eligible individuals (i.e. individuals that satisfy both the age and unemployment duration requirements). Each cell in the table shows the result of the average treatment effect on the treated (ATT) based on matching on the propensity score for the period before (column 1) and after (column 2) the program implementation. Each of these estimates is obtained using as treatment group individuals in employment offices that implemented the program and as control group registered unemployed in employment offices that did not implement the program by the time the observation is being recorded. The third column in each panel is the difference-in-difference estimate of the ATT over time. It is obtained as the difference of the ATT after the program implementation minus the ATT before the program implementation. The right panel (PEP=0) presents similar results for non-eligible individuals (those not satisfying one of the two requirements).

Our results show a negative impact, although not statistically different from zero, of the program on wages. On average eligible unemployed workers finding a job in employment offices that implemented the program had a starting wage that was on average 2.4 per cent lower than the one obtained by matched workers in employment offices not implementing the program. When compared with wages obtained in the period prior to the program implementation we see that individuals in employment offices that implemented the program used to get wages that are 3.1 per cent higher than those in offices not implementing the program. This results in a D-in-D estimate of minus 5.5 per cent for the ATT on the treated. The right panel corrects these figures for general trends observed for individuals not covered by the program and thus not affected by its impact. The D-in-D estimate for this group of workers is a 8.5 per cent increase of wage since the period after the program implementation. Overall, the DDD estimate is minus 14 percent and statistically significant. Thus, after correcting for general trends on wages and for the composition of the treated group we obtained a sizable wage loss associated with the program.

We also present results for two dimensions of individual heterogeneity, namely gender and program type. In terms of gender differences, our results point to a very large and negative impact of the program on starting re-employment wages for male workers (minus 28 per cent for the DDD matching estimate) and a close to zero impact on female wages (0.3 per cent). The main source of this difference comes from the large increase in wages obtained by noneligible unemployed after the program implementation.

The second heterogeneity dimension that we consider is the type of program the individ-

Table 5: D-in-D matching and DDD matching average wage treatment effect

	PEP=1			PEP=0			
	Before	After	**D-in-D**	Before	After	**D-in-D**	**DDD**
All	0.031	-0.024	-0.055	-0.077	0.008	0.085	-0.140
	0.044	*0.028*	*0.052*	*0.035*	*0.036*	*0.050*	*0.072*
Female	-0.036	-0.020	0.016	-0.034	-0.021	0.013	0.003
	0.055	*0.031*	*0.063*	*0.043*	*0.047*	*0.064*	*0.090*
Male	0.065	-0.018	-0.083	-0.105	0.095	0.200	-0.283
	0.073	*0.057*	*0.093*	*0.060*	*0.045*	*0.075*	*0.119*
Reage	-0.014	-0.027	-0.013	-0.094	0.029	0.123	-0.136
	0.053	*0.037*	*0.065*	*0.043*	*0.051*	*0.067*	*0.093*
Inserjovem	0.078	-0.040	-0.118	-0.035	-0.080	-0.045	-0.073
	0.063	*0.046*	*0.078*	*0.078*	*0.045*	*0.090*	*0.119*

Note: Standard errors in italic.

uals were registered in. The registration in one of the two types of program was completely dictated by the individual age, but the effectiveness of the program can vary substantially. The two programs had a negative impact on earnings of registered unemployed, but the Reage program (the one targeting older workers) register a more negative impact on wages than Inserjovem (the one targeting young individuals). The former program had a negative impact of 14 per cent on starting wages, most of it as a result of the good wage performance of individuals in Employment Offices not implementing the job search program. On the contrary, the results for the Inserjovem individuals were less negative, and statistically not different from zero, but they were mainly driven by the performance of young workers in Employment Offices in the program.

5 Conclusions

The purpose of this study has been to evaluate labor market programs for youth and long-term unemployed in Portugal using as measures of effectiveness the (re-)employment probability and wages. The programs evaluated – REAGE and InserJovem – are job-search support programs.

We identified the average treatment effect based on the hypothesis that participation in the various treatments, including the no-treatment state, is independent of the post-program outcomes conditional on observable exogenous factors (as well as non-observable factors in our DDD implementation). The mandatory and phased implementation characteristics imposed on the design of the program allows us to be confident about our identification strategy, namely the comparability of our treatment and control groups.

The results from our analysis point to a positive, but rather small, effect of the treatment on unemployment duration on the treated group. We estimate a reduction of less than 1 month on unemployment duration. Given the generally high levels of unemployment duration in Portugal (that can reach several years) these numbers are not impressive. Indeed, they are in line with what has been obtained for other countries and surveyed in Heckman et al. (1999). In terms of the other chosen measure of effectiveness of the program – wages –, qualitatetively the results are in line with the duration ones. The average re-employment wages effect on the treated women is null and negative for men. Overall, we conclude that the program effectiveness can be questioned.

References

Becker, S. O. & Ichino, A. (2002), 'Estimation of average treatment effects based on propensity scores', *The Stata Journal* **2(4)**, 358–377.

Blundell, R. & Dias, M. (2000), 'Evaluation methods for non-experimental data', *Fiscal Studies* **March**.

Blundell, R., Dias, M., Meghir, C. & Reenen, J. V. (2004), 'Evaluating the employment impact of a mandatory job search assistance program', *Journal of the European Economic Association* **forthcoming**.

Heckman, J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**, 153–161.

Heckman, J., Ichimura, H., Smith, J. & Todd, P. (1998), 'Characterizing selection bias using experimental data', *Econometrica* **66(5)**, 1017–1098.

Heckman, J., Ichimura, H. & Todd, P. (1997), 'Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme', *The Review of Economic Studies* **64(4)**, 605–654.

Heckman, J., Ichimura, H. & Todd, P. (1998), 'Matching as an econometric evaluation estimator', *The Review of Economic Studies* **65(2)**, 261–294.

Heckman, J., LaLonde, R. & Smith, J. (1999), 'The economics and econometrics of active labor market programs in', *Handbook of Labor Economics* **3A**.

Larsson, L. (2003), 'Evaluation of swedish youth labor market programs', *The Journal of Human Resources* **38(4)**, 891–927.

Meyer, B. D. (1995), ‘Natural and quasi-experiments in economics’, *Journal of Business & Economic Statistics* **13**, 151–162.

Rosenbaum, P. & Rubin, D. (1983), ‘The central role of the propensity score in observational studies for causal effects’, *Biometrika* **70**, 41–55.

Rubin, D. (1977), ‘Assignment to a treatment group on the basis of a covariate’, *Journal of Educational Statistics* **2**, 1–26.

Smith, J. & Todd, P. (2004), ‘Does matching overcome LaLonde’s critique of nonexperimental estimators?’, *Journal of Econometrics* **forthcoming**.