



Evaluating job-search programs for old and young individuals: Heterogeneous impact on unemployment duration

Luis Centeno^a, Mário Centeno^b, Álvaro A. Novo^{c,*}

^a ISCTE/INDEG Business School, Portugal

^b Banco de Portugal & ISEG-U. Técnica de Lisboa, Portugal

^c Banco de Portugal & ISEG-U. Nova & U. Lusófona, Portugal

ARTICLE INFO

Article history:

Received 15 April 2005

Received in revised form 22 November 2007

Accepted 22 February 2008

Available online 18 March 2008

JEL classification:

J18

J23

J38

Keywords:

Active labor market program

Causal evaluation

Difference-in-differences estimators

Unemployment duration

ABSTRACT

This paper exploits an area-based pilot experiment to identify average treatment effects on unemployment duration of treated individuals of two active labor market programs implemented in Portugal. We focus on the short-term heterogeneous impact on two subpopulations of unemployed individuals: young (targeted by the Inserjovem program) and old (targeted by the Reage program). We show that the latter program has a small and positive impact (reduction) on unemployment duration of workers finding a job upon participation, whereas the impact of Inserjovem is generally negative (extended durations). These results are robust to a wide variety of constructions of quasi-experimental settings and estimators. The identification of heterogeneous effects showed that the program results were less satisfactory for young workers, for those over 40 and for the less educated. Women also benefited less from the programs. The results seem to improve slightly for young workers in the 2nd semester of implementation, but they deteriorate in the medium term. The lack of wage subsidies in the Portuguese programs may explain the minor impacts obtained, when compared to similar programs.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The effectiveness of active labor market programs (ALMP) in reducing unemployment and speeding transitions into employment has been the subject of a large and growing literature, as reviewed in Heckman et al. (1999) and Kluve (2006). We evaluate a set of ALMP directed at both old and young Portuguese unemployed. In the late 1990s, Portugal developed two initiatives – Reage and Inserjovem – aiming at increasing the employability of the long-term unemployed (the Reage program), and acting earlier on youth unemployment, preventing episodes of long-term unemployment at the beginning of their labor market careers (the Inserjovem program).¹ This emphasis on preventive actions on long-term unemployment led us to choose the impact on unemployment duration as the outcome of interest in our evaluation exercise.

The most distinctive feature of our study is the possibility of evaluating the impact of ALMP on young and older workers at the same time. In recent years, employment policies have started paying further

attention to older individuals (OECD, 2006). The promotion of longer working lives due to issues of population ageing and pension system funding brings about increasing challenges to ALMP. A larger pool of older unemployed requires the attention of public policies, but also the challenges are different from those faced by younger unemployed. Traditionally, the older group faces harder labor market conditions; Addison et al. (2004) report evidence of a non-stationary labor market environment, namely, that the arrival rate of job offers falls sharply with age. Additionally, past experience with ALMP shows that older workers are less motivated to participate in some of the initiatives (e.g. training, a key component in the learning for life approach). By focusing on the differentiate impact across age groups, this paper contributes to this strand of the literature.

The relevance of our study is further potentiated by the institutional rigidity of the Portuguese labor market. Despite being applied during a period of low unemployment, the Portuguese labor market is characterized by extremely high employment protection, long unemployment spells and generous unemployment insurance, and a low arrival rate of job offers, even for European standards. Overall, the Portuguese setting constitutes a challenging environment for any ALMP. Thus, in the context of the European Employment Strategy, a rigorous evaluation of the Portuguese experience with ALMP applied to such a diverse group of individuals may be of great relevance for

* Corresponding author.

E-mail addresses: luisgcenteno@gmail.com (L. Centeno), mcenteno@bportugal.pt (M. Centeno), anovo@bportugal.pt (Á.A. Novo).

¹ We will refer to Reage and Inserjovem as ‘programs’, although, as it is common with ALMP in place in most European countries, they include a large set of policies to be detailed below.

the implementation of similar programs in other countries. Indeed, we find evidence of an heterogeneous impact of the programs. For individuals exiting into employment, those aged 30–40 reduced the duration of their unemployment spells, while for older individuals (over 40) and young cohorts (less than 25) the programs had no impact.

The programs under evaluation target all young people (less than 25 years old) before they have been registered for 6 months and all adults before they reach 12 months of unemployment. The early intervention is meant to ensure the timely implementation of responses suitable to each individual's situation. These responses are essentially job-search assistance, including vocational guidance, counseling, monitoring, and training or re-training options. Furthermore, they have a mandatory character, in the sense that failing to comply with the directions of the Employment Office (EO) placement team will result in the loss of subsidies (including unemployment insurance and fee-exemption to access the public health services).

The implementation of the Portuguese programs created an almost natural setting for the evaluation exercise. The programs were first introduced in a subset of EOs beginning in June 1998, generating an area-based pilot experiment that we explore in our identification strategy of the programs' impact. Afterwards, they were rolled out sequentially to the other EOs, fully covering the country in January 2001. The pilot EOs were not randomly chosen, but neither was participation based on specific local labor market conditions. Indeed, they were picked for logistical reasons unrelated with the programs' goals in terms of labor market outcomes. We apply a difference-in-differences methodology using the natural treatment and control groups originating from this pilot setting. The treatment group will consist of all registered individuals who participated in the programs in pilot EOs and the counterfactual is drawn from the subset of EOs not implementing the program in the evaluation period. The difference-in-differences methodology is supplemented with a combination of matching methods to generate the difference-in-differences matching estimator, used to eliminate some potential sources of bias Heckman et al. (1997).

The goal of this paper is to determine the effects of the programs compared to the outcome of the individual had (s)he continued to search for a job in the absence of the support provided by the programs. We focus on the duration of complete spells of unemployment of individuals leaving the programs during the first 12 months of implementation (a short-term evaluation) and 2 years later (a medium-term evaluation). Given the wide coverage and mandatory nature of the programs, they can be considered as having comprehensive implementation. This raises the possibility of observing indirect effects, such as substitution and equilibrium effects. The first effect operates through a change in the relative price of labor between the treated and untreated individuals and the second through an increase in the supply of labor that lowers wage and increases employment. The evidence collected by Katz (1998) shows that these effects are more likely to arise in the context of wage subsidies, which were not included as a treatment in the Portuguese programs, contrary to what has happened in other countries, for example, the U.K. and Sweden. In our empirical work, we present some evidence on the dimension of possible substitution effects by exploring the regional implementation of the programs.

Previous microeconomic studies of ALMP in European countries, taking place at around the same time period, find mixed results. Blundell et al. (2004) find an important "program introduction effect" for the UK; the program effect is much larger in the first quarter than later on. These results are confirmed in the longer-run analysis of De Giorgi (2005) for a sample of young males, using a regression discontinuity approach; he finds no evidence of substitution or general equilibrium effects. Larsson (2003) and Sianesi (2004) find no significant effect in the Swedish programs; the effects are small and positive for the employment rate, but negative on reemployment

wages. Still for Sweden, Carling and Richardson (2004) conclude that transition rates improve more for subsidized work experience and training provided by firms than those observed for classroom vocational training. For East Germany, Eichler and Lechner (2002) find that participation in public employment policies implemented after reunification reduces participants' probability of employment. For a massive ALMP implemented in the late 1990s in Switzerland, Gerfin and Lechner (2002)'s evaluation finds a positive impact for a policy involving a wage subsidy, but negative effects for traditional employment programs.

An alternative strand of the literature conducts the evaluation of ALMP with duration models, such as in the work of Bonnal et al. (1997) for France, Eberwein et al. (2002) for the US Job Training Partnership Act and van den Berg and Klaauw (2006), who study ALMP applied in the Dutch labor market. The minor impacts on the labor market prospect of the unemployed involved in ALMP are typically confirmed in these alternative studies.

Our assessment of the Portuguese programs points to a small reduction in unemployment duration. In the absence of the program, we estimate that unemployment duration of treated individuals would increase by at most 0.4 of a month, which would not represent a large increase in duration given that some workers spend many months unemployed. The results show some degree of heterogeneity for different types of exits and programs. Recipients of treatment in *Inserjovem* tend to benefit more when moving into training (a reduction in unemployment duration) than into employment (with an increase in unemployment duration), although the impact never exceeds 0.3 of a month. For older workers, *Reage* program, the results point in the other direction with shorter unemployment spells for individuals placed in a job upon program participation and longer spells if they enter training.

Gender, age and schooling seem to play an important role in determining the programs' impact. In transitions into employment, the impact is larger for men in *Reage* (a reduction of unemployment close to one month). In terms of age, the largest impact is observed for individuals aged between 30 and 40. Workers with a higher degree of education seem to benefit more from the programs, especially those in *Reage*. When the programs are evaluated for cohorts that were treated after the initial implementation phase, the results point towards similar impacts in the second semester of implementation, but to longer unemployment spells when evaluating the cohorts two years after June 1998; a pattern similar, for example, to Blundell et al. (2004).

The results are robust to the choice of treatment and control groups and we were not able to find clear spillover effects from treated to untreated areas in terms of unemployment duration.

The modest results of the Portuguese programs in reducing unemployment duration might be explained by the lack of some key treatments, such as wage subsidies that, as claimed in Katz (1998), increase ALMP effectiveness for specific disadvantaged groups. In the context of the "live longer, work longer" debate, the paucity of the results for older workers (aged more than 40) points towards the importance of designing policies specific to these workers and, more generally, to the educationally disadvantaged groups.

The paper is organized as follows. The labor market programs are described in Section 2. The evaluation problem, as well as the identification and estimation of average treatment effects are addressed in Section 3. Section 4 presents the data and the results are discussed in Section 5. Finally, concluding remarks are presented in Section 6.

2. The programs

2.1. Description of *Inserjovem* and *Reage* programs

We study large-scale ALMP implemented in Portugal at the end of 1990s, in the context of Guidelines I and II of the European Employment Strategy, which aimed at improving employability and developing

entrepreneurship. Similar programs implemented in other European countries have been subjected to evaluation (see, for example, Larsson, 2003 or Sianesi, 2004, for studies of the Swedish Youth Practice Program, and Blundell et al., 2004 or De Giorgi, 2005 for the British New Deal Program; Kluve, 2006 surveys microeconomic evaluation of ALMP in European countries). The Portuguese programs aimed at preventing long-term unemployment and are primarily composed of job-search support initiatives, involving vocational guidance, counseling, monitoring and training. The main policy goal was to improve the employability of two target groups: (i) individuals aged less than 25 years, who must be enrolled in the program prior to completion of six months of unemployment (the Inserjovem program) and (ii) individuals aged 25 or more, who must be enrolled before the 12th month of registered unemployment (the Reage program).²

Program participation is mandatory; all eligible individuals who refuse to participate face a loss of entitlement to benefits and their registration is cancelled. The benefits of being registered at the EO are not confined to unemployment insurance, but include fee-exemption to access the public health services, and other programs, such as training schemes, offered by the Employment Services. The mandatory characteristic of the programs reduces the potential for self-selection into the programs, which constitutes the main source of bias in other evaluation exercises (Michalopoulos et al., 2004).

The programs are composed of intensive job-search assistance and small basic skills training, for example, writing a curriculum vitae. They include a large number of different responses by the EO placement team. Each individual is interviewed with placement officers to help her/him improve job-search skills and, if deemed necessary, (s)he can enter a number of vocational or non-vocational training courses. The whole process of job-search assistance ends with the elaboration of a “Personal Employment Plan”, which includes detailed information on the unemployed individual's job-search effort. According to this Plan, the unemployed individual is expected to meet on a regular basis with the placement officer and to actively search for a job. Unjustified rejection of job offers leads to the cancellation of registration. The programs led EOs to pay a much closer attention to the search commitment of registered unemployed.

Individuals enrolled in the programs face different exits. The main ones, upon participation in the programs, are placement in a salaried job available through vacancies posted in the EOs, and education or training.

The programs were launched in June 1998 in a limited number of EOs. Afterwards, they were gradually extended to the rest of the country until full coverage was reached in January 2001.³ The EOs selected to participate in June 1998 generated an area-based pilot group that we explore for the evaluation exercise as our treatment group. As it can be seen in Fig. 1, these EOs were evenly spread around the country, with the possible exception of the southern region of Alentejo. Additionally, these offices were not chosen because of particular labor market conditions prevailing at that moment. Instead, they were selected because they were logistically ready to comply with the technical requirements of implementing the programs. All other EOs are potential control units. However, since our main results focus on a six months' time window, we had to exclude as control units all EOs that joined the programs later on, but within the time window, namely those joining in October, 1998. The municipalities belonging to each of these 3 groups (treatment, control and excluded) are identified in Fig. 1.

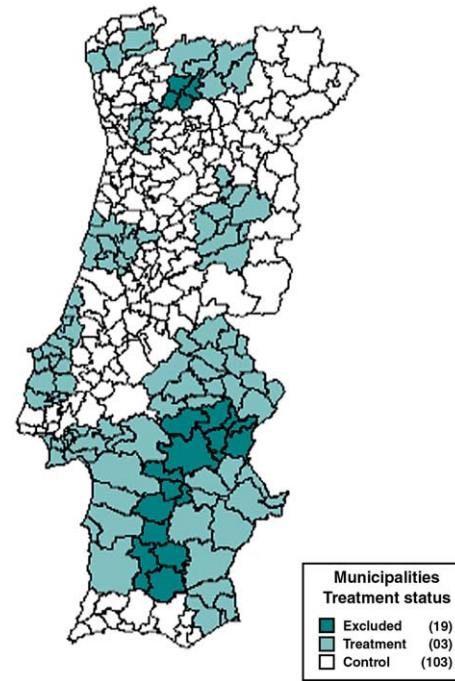


Fig. 1. Mainland Portugal: Treatment and control employment offices (EO) geographical location (number of municipalities in parentheses); each EO may cover more than one municipality.

Between June 1998 and December 2002 there were about 1.5 million Portuguese unemployed individuals registered in EOs across the country, of which roughly 61% were women and 35% were young (under 25). These numbers give an idea of the dimension and wide coverage of the programs. The implementation of these programs also implied a significant increase in the volume of expenditure. Indeed, between 1997 and 1999, comparing a pre-program period and the pilot period (i.e., not yet covering the whole country), the outlays specific to the ALMP increased by 60%, to close to 90 million euros (these figures do not include the costs shared by all EOs services).

2.2. Outcome of interest

Active job-search programs are aimed at easing/speeding the transition from unemployment to employment. This study seeks to evaluate the impact of two programs on the duration of unemployment spells on each of the two targeted groups – young individuals experiencing short-lived unemployment (under 25 and unemployment spells above 6 months) and older workers with longer unemployment spells (25 years old or more and spells over 12 months). The parameter of interest in our exercise is the impact on the average unemployment duration among the treated as a result of improved job-search assistance.

We will measure the duration of unemployment spells that, after a program intervention, terminate in one of the following states: (i) employment, (ii) training, or (iii) cancelled registrations. In the following section we present the identification issues raised in the selection of the control group and the estimation methods used.

3. Identification and estimation methods

The problem of evaluating active labor market programs has been extensively studied in the literature (Heckman et al., 1999). The methods draw on the use of natural experiments to evaluate treatment effects in the absence of truly experimental data. These methods suggest several solutions to the problem of generating a conveniently designed comparison group necessary to perform the program

² In practice, at each pilot EO, there was a stock of youth (older) unemployed registered for at least 6 (12) months, who, according to the officials involved in the implementation of the programs, were also quickly enrolled.

³ The implementation dates are: June and October, 1998; February, May, July, and November, 1999; April, June, and September, 2000; and January, 2001.

evaluation. Given the quasi-experimental feature of these programs, the feasibility of any evaluation exercise depends crucially on the ability that researchers have to generate such counterfactual groups from the data available on the program implementation.

Typical methodologies proposed to tackle these issues include: difference-in-differences (see Meyer, 1995 and the recent review in Abadie, 2005) and matching methods (Rubin, 1977; Rosenbaum and Rubin, 1983; Rubin, 2006). The difference-in-differences matching method has been proposed by Heckman et al. (1997) and Heckman et al. (1998) as a combination of the two former methods. It was recently reviewed and compared with the other methods by Smith and Todd (2005) and has the potential benefit of eliminating some sources of bias present in quasi-experimental settings, improving the quality of evaluation results significantly.

We take advantage of the characteristics of the dataset and of the program implementation to construct treatment and control groups. In particular, we explore (i) the existence of data for the pre- and post-program periods, and (ii) the source of variation resulting from the pilot implementation phase (which generates spatial and time differences). The available sample has pre-treatment and post-treatment observations and we use this repeated cross-section characteristic to implement estimators from the difference-in-differences class. The dataset characteristics are also important in trying to meet the fundamental requirements for the success of our evaluation exercise, namely, choosing a comparison group from the same local labor market and with comparable measures from a common data source, as put forward by Heckman et al. (1997) and Heckman et al. (1998).⁴ The second requirement is clearly met in our evaluation exercise as all information was recorded through a common dataset obtained from the administrative records of the EOs' database (see Section 4). Next we concentrate on the first argument, usually identified as the main source of bias in evaluating mandatory public programs (Michalopoulos et al., 2004).

3.1. The treatment groups and the counterfactuals

The design and phased implementation of the programs across Portugal generated a natural way to construct treatment and control groups. The treated individuals are drawn from EOs that implemented the programs in June 1998 (i.e., the pilot EOs). In particular, the Inserjovem treatment group includes all individuals under 25, who completed an unemployment spell of 6 months either before June 1998 or during the first 6 months of implementation; the Reage treatment group is defined in the same fashion, but it consists of individuals 25 years old or more with at least 12 months of unemployment.

The construction of the comparison group was determined by the same criteria, but considering, instead, EOs not offering the programs during the period of evaluation.

This paper examines the two age groups separately: the young in Inserjovem and the older in Reage. Individuals, within each of these groups, will probably be fairly homogeneous across treatment and comparison sites and the mandatory nature of the program reduces the potential of self-selection into the program to produce a large bias. Thus, the main source of bias in these comparisons might be geographical mismatch, something we will deal with carefully in our empirical exercise.

As mentioned, the timing of implementation at each EO was not random, and this raises some concerns associated with the quasi-experimental nature of the process. However, the sequencing of enrollment of each EO was not dictated by the specific labor market conditions prevailing at the regional level, and this allows us to argue that it does not bias our estimates. As previously shown, they are spread across the country (Fig. 1) and, as it can be seen in Fig. 2, except

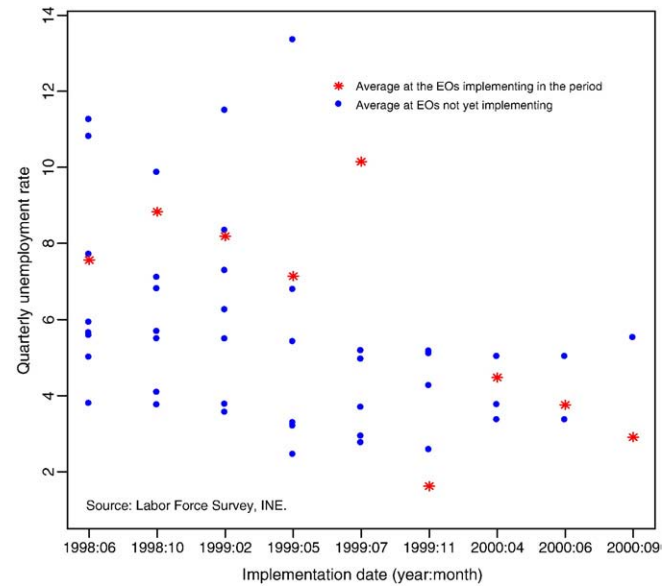


Fig. 2. Employment Offices (EOs) quarterly average unemployment rate at the implementation dates. Notes: The figure lists the sequence of EOs enrollment dates, starting in June 1998 up to September 2000. Each data point represents a set of EOs defined by their specific enrollment date. The starred point at each date identifies the set of EOs enrolling at that specific date.

for one date, the regions with the highest (or lowest) unemployment rate were not those to be selected first to participate in the programs. Furthermore, the unemployment rate prevailing at the locations initially applying the programs is close to the average unemployment rate prevailing at other locations.

We understand that the conditions to evaluate the impact of the policy in such a setting are not as perfect as in experimental settings and, as such, identification of the treatment effect requires stronger conditions than if these EOs had been assigned in a truly random fashion. In fact, in our exercise, the counterfactual is drawn from a different labor market. We will take this into account when discussing identification, and will try to circumvent the possible sources of bias still remaining in the quasi-experimental setting through the choice of adequate estimation methods. We will also do several robustness checks of our identifying assumptions.

3.2. Econometric implementation

Let Y_{it}^D be the outcome of interest for individual i at time t given that (s)he is in state D , where $D=1$ if exposed to the program, and 0 otherwise. Let treatment take place at time t . The fundamental identification problem lies in the fact that we do not observe, at time t , individual i in both states. Therefore, we cannot compute the individual treatment effect, $Y_{it}^1 - Y_{it}^0$. It is possible, however, if provided with a convenient control group, to estimate the average effect of the treatment on the treated. The method we use is often labelled difference-in-differences (D-in-D) and compares the average behavior before and after the program for the treatment group with the before and after outcomes for the comparison group (Blundell and Costa Dias, 2000).

The idea behind a D-in-D estimator is that we can use an untreated comparison group to identify temporal variation in the outcome that is not due to the treatment. However, in order to achieve identification of the general D-in-D estimator we need to assume that the average outcomes for treated and controls would have followed parallel paths over time. This is known as the common trend assumption:

$$E[Y_{it}^0 - Y_{it}^1 | D = 1] = E[Y_{it}^0 - Y_{it}^1 | D = 0], \quad (1)$$

⁴ The availability of a rich set of variables related to program participation is also indicated as crucial for the good performance of some of the methods used in the estimation, but we will refer, at full length, to this aspect later.

where t' is a time period before program implementation. The assumption states that the temporal evolution of the outcome variable of treated individuals ($D=1$), if they had not been exposed to the treatment, would have been the same as the observed evolution for the individuals not exposed to the treatment ($D=0$).

If the assumption expressed in Eq. (1) holds, the D-in-D estimate of the average treatment effect on the treated can be obtained by the sample analogs of

$$\alpha_{D-in-D} = \{E[Y_{it}|D=1] - E[Y_{it}|D=0]\} - \{E[Y_{it'}|D=1] - E[Y_{it'}|D=0]\}, \quad (2)$$

where Y_{it} is the observed outcome for individual i at time t . As stated above, it measures the impact of the program by the difference between participants and non-participants in the before-after difference in outcomes.

The common trend assumption can be too stringent if the treated and control groups are not balanced in covariates that are believed to be associated with the outcome variable (a common problem referred to as the Ashenfelter's dip, after Ashenfelter, 1978). The D-in-D setup can be extended to accommodate a set of covariates and this is usually done in a linear way. The formulation below takes into account eligibility specific effects and time/aggregate effects. In the following model, α_D corresponds to this estimate and is obtained for a sample of treatment and control observations:

$$Y_{it} = \lambda D + \eta \tau_t + \theta Z_{it} + \alpha_D D \tau_t + \varepsilon_{it}, \quad (3)$$

where D is as before and represents the eligibility specific intercept, in our case defined over age, unemployment duration and area according to each of the programs' rules, τ_t captures time/aggregate effects and equals 0 for the pre-program period and 1 for the post-program period. Z is a vector of covariates (pre-determined with respect to the introduction of the program) included to correct for differences in observed characteristics between individuals in treatment and control groups.

This estimator allows us to control for differences in the Z and allows for time specific effects but it does not allow α_D to depend on Z and it does not impose common support on the distribution of the Z 's across the cells (before and after; treatment and control) defined by the D-in-D approach. Additionally, and as pointed out by Meyer (1995), this procedure might be inappropriate if the treatment has different effects for different groups in the population.⁵

These pitfalls can be mitigated by supplementing the D-in-D estimates with propensity score matching. The difference-in-differences matching (DDM) estimator implemented follows Heckman et al. (1997, 1998), and Smith and Todd (2005). Intuitively, the benefits over the simple D-in-D estimator may arise from the fact that the matching version adds the comparability on the observable covariates that characterizes the propensity score matching estimator.

The feasibility of the matching strategy relies on a rich set of observable individual characteristics, those included in Z , and we use it to guarantee that the distribution of the individual characteristics important for the definition of the outcome is the same in the difference-in-differences cells. The matching process models the probability of participation and matches individuals with similar propensity scores. The common trend assumption behind the DDM estimator (Smith and Todd, 2005) is

$$E[Y_{it}^0 - Y_{it'}^0|P(Z), D=1] = E[Y_{it}^0 - Y_{it'}^0|P(Z), D=0], \quad (4)$$

where $P(Z) = \Pr(D=1|Z)$ is the propensity score. The set of variables in Z typically includes information on the demographic characteristics

and recent labor market history of workers. Thus, the key identifying conditions on growth is conditional on Z , and via properties of the propensity score (Rosenbaum and Rubin, 1983) this implies Eq. (4). The DDM estimator also requires the support condition of a non-participant analogue for each participant, formally that $\Pr(D=1|Z) < 1$. As pointed out in Heckman et al. (1997, 1998), it is important to guarantee that we are able to find a match for program participants, that is, that the condition $\Pr(D=1|Z) < 1$ is satisfied. Indeed, if there are regions where the support of Z does not overlap for the treatment and control groups, we should restrict the matching to the common support region to avoid a major source of bias in quasi-experimental program evaluations (Heckman et al., 1998).

We use the repeated cross-sectional dimension of our data, and implement the DDM estimator following Smith and Todd (2005):

$$\hat{\alpha}_{DDM} = E[Y_t^1 - \hat{E}(Y_t^0|P)] - E[Y_{t'}^1 - \hat{E}(Y_{t'}^0|P)], \quad (5)$$

where $\hat{E}(Y_t^0|P)$ and $\hat{E}(Y_{t'}^0|P)$ represent the expected outcome of individuals in the control group matched with those in the treatment group, respectively, in the after and before periods. To guarantee that, in each time period, the distribution of the relevant characteristics is the same for treatment and control, we ensure that observations in the four cells (Before/After; Treatment/Control) have propensity scores within the same range. In practical terms, we restrict the computation of the average treatment effect on the treated to units that, after being successfully matched in each period, have propensity scores within a common range among the four cells, ensuring common support as required in Smith and Todd (2005).⁶

In each time period, units are matched using both kernel and spline matching algorithms.⁷ The implemented algorithm is described in Section 1 of the Appendix A.

4. Data

The Portuguese EOs collected data on the entire population of registered unemployed individuals regardless of their participation status in the programs. The SIGAE dataset covers both the pre- and post-program periods. Our data start in January 1997, 1.5 years before the start of the programs, and finish in December 2002, almost 2 years after the last set of EOs joined in. Overall, the dataset comprises over 2 million spells of registered unemployment, monitoring different features of the programs and individuals. The information in the dataset includes most demographic variables used in labor market studies (e.g. age, sex, nationality, schooling, place of residence) and some variables related with previous labor market experience (e.g. reason for job displacement). The unemployed individual is observed for the 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).⁸

In the empirical analysis, the post-program period covers the 6 months time window from July to December 1998. The pre-program period covers the 6 months preceding the policy implementation (December 1997 to May 1998). In analyzing duration data, the statistical description of the process depends both on the transition intensities and on the way the data is collected. In our case, we use a sampling plan in which we take from the whole population of registered unemployed all individuals leaving unemployment in the mentioned fixed time window. Following Lancaster (1992, p. 182), the

⁶ A variant of the DDM estimator for repeated cross-sections data appears in Blundell et al. (2004).

⁷ See Becker and Ichino (2002) and Caliendo (2006) for a practical description of these and other matching criteria. The empirical implementation is partially based on Leuven and Sianesi (2003).

⁸ A follow-up interview with program participants was initially considered but the EOs were not able to carry them out in a systematic way, and the interview process was eventually abandoned.

⁵ Note that heterogeneity in the treatment effect can be accommodated in Eq. (3) by suitably interacting D with the Z 's (see Heckman and Hotz, 1989).

Table 1
Summary statistics of pre-determined characteristics: by program and treatment status

Variables	Reage			Inserjovem		
	Mean		Difference	Mean		Difference
	Control	Treatment		Control	Treatment	
Age	41.02	41.34	0.33	21.36	21.18	−0.18
Female	0.55	0.59	0.05	0.62	0.64	0.02
Married	0.72	0.74	0.02	0.15	0.16	0.00
Handicapped	0.01	0.01	0.00	0.01	0.01	0.00
Foreigner	0.01	0.01	0.00	0.01	0.01	0.00
Reason to register						
Inactive	0.04	0.05	0.01	0.38	0.41	0.02
Quitted	0.05	0.05	0.00	0.05	0.04	−0.01
Fired ⁽¹⁾	0.64	0.60	−0.04	0.41	0.37	−0.04
Other	0.27	0.30	0.03	0.16	0.18	0.02
Education level						
4 or less years	0.51	0.56	0.05	0.09	0.09	0.00
6 years	0.20	0.20	0.00	0.29	0.29	0.00
9 years	0.13	0.12	−0.01	0.23	0.25	0.02
High school	0.11	0.09	−0.02	0.30	0.31	0.01
College	0.05	0.03	−0.02	0.09	0.07	−0.03
Regions						
North	0.49	0.16	−0.33	0.50	0.18	−0.32
Center	0.10	0.19	0.09	0.14	0.21	0.07
Greater Lisbon	0.38	0.43	0.05	0.32	0.40	0.07
Alentejo	0.00	0.17	0.17	0.00	0.16	0.16
Algarve	0.03	0.02	−0.01	0.03	0.02	−0.01
Unemployment spell						
Complete durations ⁽²⁾	32.39	30.48	−1.91	14.81	14.11	−0.69
No. observations	31,121	11,285		24,511	10,879	
By unemployment exits						
Placed ⁽³⁾	6037	2424		3950	1769	
Training	694	482		1017	674	
Cancelled ⁽⁴⁾	24,390	8379		19,544	8436	

Notes: The Reage program targeted 25 years old or more with unemployment spells of at least 12 months. The Inserjovem targeted 15 to 24 years old with unemployment spells of at least 6 months. Treatment includes eligible individuals in pilot employment offices for the July to December 1998 period. The control group includes eligible individuals in non-pilot offices. The data are collected by the Employment Services and recorded in the SIGAE database. (1) It includes end of temporary job and fired unemployed. (2) Unemployment spell duration measured until the individual exits unemployment. (3) Refers to transitions to employment using EOs services or own placement. (4) Refers to registrations cancelled by the EOs due to irregularities.

analysis of the registered unemployment duration process can simply be carried out “either by observing entrants to a state and their x 's [characteristics] and waiting until t [is] revealed or by observing leavers and asking from x , [and] t .” We follow the latter sampling scheme. Its main feature, under the standard stationarity assumption in this literature, is that the probability that a person is selected for the sample does not depend upon time (t). Thus, our analysis will be based only on the distribution of complete spells of unemployment leavers.

Table 1 documents pre-program period average characteristics of unmatched treated and control groups for both programs. The subjects in the treatment and control groups present already at the outset rather similar average characteristics, this is perhaps not surprisingly given (i) the common source for the data, (ii) the almost random selection process of EOs into the pilot group (see Section 1) and (iii) the proximity of the labor markets in each group (see Fig. 1). It is also worth noticing that since the dataset covers the whole population of registered unemployed, we are able to obtain a rather large sample size, the exception being the sample of transitions into training, a less common way out of unemployment.

This evidence constitutes a good stepping stone from which to proceed more confidently with the econometric analysis, i.e., we have a good starting point which together with the battery of formal tests to be conducted and the institutional setting will help us provide a convincing argument to the effect that the hypotheses needed to

implement the matching estimator, e.g. the conditional mean independence assumption, are likely to hold in our setting.

5. Unemployment duration: average treatment effect on the treated

We now proceed with the empirical evaluation of the impact of the Reage and Inserjovem programs. The unemployed can transit out of unemployment either into employment or into inactivity in the form of training or cancelled registrations. To start with, we expose the data to different methods of estimation of the treatment effect to assess the sensitivity to methodological issues. Then we consider the possibility of heterogeneous effects among treated individuals. Finally, we assess the robustness of our results by considering alternative constructions of our evaluation sample and test for some form of anticipation, timing (medium term), and substitution effects between treated and untreated individuals in different geographical locations.

5.1. A smorgasbord of estimators

In the evaluation literature there is not a *one-size fits all* methodology. On the contrary, as several authors have shown, starting with the seminal work of LaLonde (1986), but including also the studies of Dehejia and Wahba (1999, 2002) and Smith and Todd (2005), the choice of a method of evaluation is clearly dependent on the particular evaluation setting. As we have argued, the implementation of the Portuguese ALMP followed closely a natural experiment. As such, a simple difference of sample averages of treatment and control groups, in the pre- and post-program periods, might provide one first (good) guess at the impact of the programs. However, the ‘quasi’ characteristic of the experiment means that confounding factors may exist and these must be taken into consideration in the evaluation. Given the mandatory nature of the programs, we do not expect issues of self-selection into treatment. However, there might be differences in observable characteristics, driving, in particular, the outcome and not so much the selection into treatment. Thus, in Table 2, we present the results of our evaluation using a smorgasbord of estimators proposed in the literature that attempt to handle such confounding factors. In

Table 2

Short-term average treatment effect on the treated (in months): by program and destination state out of unemployment

Exit type	Program	No. observations	D-in-D		D-in-D matching ¹	
			Unrestricted ²	Restricted ³	Kernel	Spline
			(1)	(2)	(3)	(4)
Placed	Reage	18,323	−0.39 (0.44)	−0.19 (0.42)	−0.40 (0.44)	−0.36 (0.46)
	Inserjovem	11,944	−0.11 (0.34)	0.03 (0.33)	0.29 (0.36)	0.27 (0.33)
Training	Reage	2182	0.95 (1.22)	1.19 (1.20)	1.70 (1.59)	1.29 (1.63)
	Inserjovem	3168	−0.98 (0.62)	−0.91 (0.61)	−0.21 (0.78)	−0.16 (0.80)
Cancelled	Reage	68,960	−0.03 (0.25)	0.28 (0.23)	0.13 (0.28)	0.21 (0.29)
	Inserjovem	55,595	−0.54 (0.17)	−0.54 (0.16)	−0.26 (0.18)	−0.38 (0.18)

Notes: (1) The variables included in the estimation of each of the propensity scores upon which matching is performed are listed in the Appendix A, Table A1. For the case of individuals placed in the labor market, the results of the balancing property are also reported in the Appendix A, Table A2. Standard errors are presented in parentheses. For the D-in-D Matching estimates the standard errors are bootstrapped-based, using 50 replications for each estimate; (2) The column “Unrestricted” refers to the D-in-D estimator based only on the simple differences of sample averages of the dependent variables. That is, no attempt to control for observable characteristics is made; (3) The column “Restricted” refers to the D-in-D estimate obtained by estimating Eq. (3) with linear regression, controlling for the individuals’ observed characteristics listed in Table A.1.

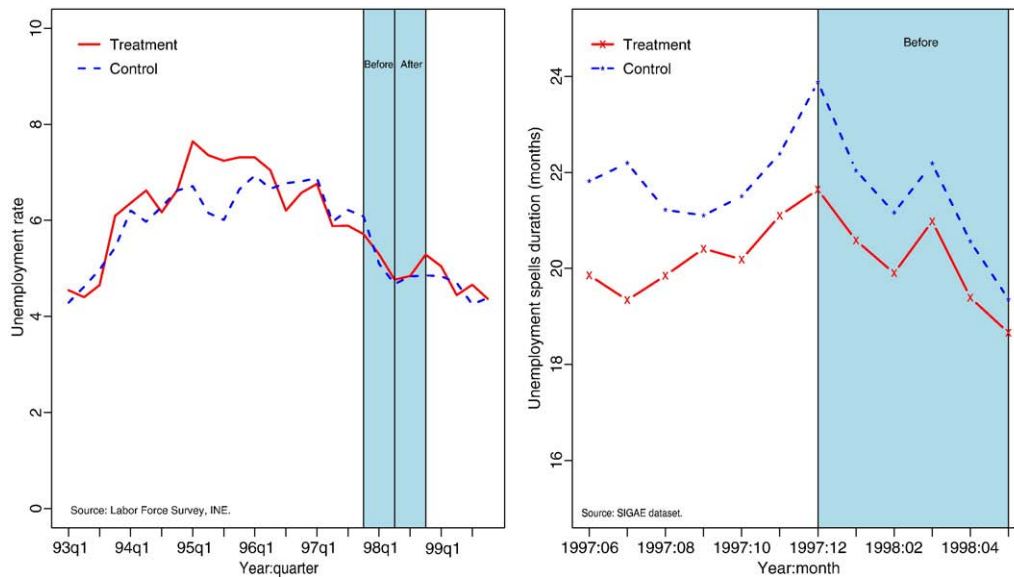


Fig. 3. Unemployment rate (left-panel) and unemployment duration (right-panel). Notes: Quarterly average unemployment rates, 1993 to 1999, for the treatment areas and control areas (left-panel). Monthly average unemployment spells duration for treatment and control areas (right-panel). The period from 1997:12 to 1998:05 corresponds to the 'before' period in the D-in-D evaluation.

particular, we will base our estimates on the class of difference-in-differences estimators, using some of the variants suggested in the literature, each of which has different underlying hypotheses of identification.

The validity of the common trend assumption is crucial to the identification of our estimates. Thus, we are interested in selecting treatment and control areas such that, in terms of pre-program average unemployment duration, the common trend assumption is satisfied. Fig. 3 represents the recent history of average unemployment rates (left-panel) and unemployment duration in the treatment and control areas (right-panel). The figure shows that the path of unemployment rates and the average unemployment duration in the two sets of EOs are very similar, suggesting that they have been subject to macroeconomic shocks to which their labor markets responded in a similar way. It is also evident from right-panel that individuals in the control areas have longer unemployment duration. However, we are more concerned with the differences between regions over time than with the differences at a given moment in time. For the common trend assumption to be met, we would like to have a constant difference over time, which would signal that the aggregate trends affect the two groups in a similar way. In fact, as depicted in right-panel, the average unemployment duration in treatment and control areas follows parallel paths in the pre-program period.

5.1.1. Difference-in-differences results

We start with the unrestricted difference-in-differences estimator. This estimator does not control for observable differences between treatment and control groups, but does account for (constant) non-observable differences over time. The results are reported in the column (1) of Table 2, and almost regardless of the program and destination state they indicate a small reduction in unemployment duration, particularly in view of the large weight of long-term unemployment in the Portuguese labor market.

The unrestricted D-in-D estimator identifies the average treatment effect on the treated only under certain distributional assumptions (see Heckman and Robb, 1985) that due to the stringent nature are best avoided if selection to treatment can be parameterized. Thus, we control for observable differences by estimating a linear model: the difference-in-differences restricted estimator. The set of conditioning variables is listed in Table A.1 of the Appendix A, and will be discussed

at further length below under the class of matching estimators. The estimates differ from the unrestricted ones, yielding typically smaller impacts (Table 2, column (2)). In any case, the results point towards a reduction of unemployment duration of at most 2 weeks.

Given the wide coverage of the Portuguese ALMP, there is scope for heterogeneous average treatment effects on the treated, which we handle with the DDM estimator in the following section.

5.1.2. Matching implementation and results

The DDM estimates reported in Table 2 are the difference of two average treatment effect estimates obtained with the propensity score matching method, following, as described in Section 3.2, the procedure for cross-sectional data in Smith and Todd (2005); further details of the implementation are available in the Appendix A.

In the estimation of the propensity scores, we used probit models specified with most of the available observable characteristics – age, marital status, gender, education level, reason to register (fired, quit, inactive), nationality, handicapped, and region (North, Center, Greater Lisbon, Alentejo, and Algarve). Section A.2 details the propensity score estimation process for individuals who found a job (placed) (Table A.1).⁹ Section A.3 presents a plethora of measures to assess the matching quality. The conclusion to be drawn from this exercise is that we achieve acceptable levels of covariate balancing as indicated by individual *t*-tests, the reduction in the absolute bias (Rosenbaum and Rubin, 1985) and the joint significance of covariates and pseudo- R^2 before and after matching (Sianesi, 2004).

Having discussed the generosity of our matching quality, the last columns of Table 2 present the DDM estimates. Focusing on the differences between the conditioning methods, it is worth highlighting the small differences in the point estimates. Overall, the idea emerges that the DDM estimates yield a less benign evaluation of the treatment (less negative or more positive point estimates), but most importantly that our results are not driven by the choice of a particular method. Thus, to self-contain the amount of reported results, and in view of the similarity of point estimates across methods, we will henceforth focus our economic interpretation of the programs' impact on the DDM estimates.

⁹ The results for the other exits/matching procedures are available from the authors upon request. Qualitatively, they convey similar messages.

5.1.3. D-in-D matching estimates by destination state

We now proceed by interpreting the short-term (first 6 months of implementation) results of our DDM estimates, for each of the destination states — placed, training and cancelled —, analyzed for each of the programs.

5.1.3.1. Placed. The point estimates of the treatment effect on the treated in Table 2 indicate negligible impacts of the two programs for those finding a job. Older individuals (Reage) were placed in a job about 2/5 of a month earlier than if they had not been exposed to the program, while younger unemployed (Inserjovem) saw their unemployment spells aggravated by about one week. However, both estimates are not statistically different from zero and, therefore, not different from each other. Thus, in what may be argued is the most important, but also the most challenging task of the programs, namely matching workers with jobs, the current assessment shows that the programs did not have the expected effects.¹⁰

5.1.3.2. Training. Short of being able to place the unemployed in new jobs, the employment agency may consider training as an alternative. The expected impact of programs such as Reage and Inserjovem is ambiguous. On the one hand, if they result in earlier assessment of the unemployed's needs, then such programs may bring about a reduction in unemployment spells. On the other hand, if earlier program efforts are directed towards job search, then the decision to give training may be delayed, resulting subsequently in longer unemployment periods. Furthermore, there is the possibility of locking in effects. The evidence that we collect differs, yet again, by program. This time the impact is positive (reduced duration) for youth unemployed and negative (increased duration) for the older individuals. Also notice the larger magnitude in the latter case, 1.7 to 1.3 months and the economically insignificant impact of around minus 1/5 of a month for the former group. From a statistical point of view, both estimates are not significant and with larger standard errors than in the previous estimates, which may in part be explained by smaller sample sizes (18,323 vs 2182 observations in Reage and 11,944 vs 3168 in Inserjovem). These sample sizes suggest that most of the training seems to have targeted younger individuals, even though overall Reage covers a larger portion of treated (and control) individuals. If we take the above argument for the expected impact, it is possible to maintain that there is some complementarity between job placement and training. When the program reduces unemployment spells via job placement, it increases unemployment spells before transiting into training (e.g. -0.36 vs 1.29) and vice-versa (e.g. 0.27 vs -0.16).

5.1.3.3. Cancelled. Finally, we study the impact that the programs might have had on raising awareness of irregularities that may lead to the cancellation of registrations. Again, the impact is different for the two programs. In the Inserjovem program, the average treatment effect on the treated is positive and statistically significant: a reduction in the unemployment period of up to 2/5 of a month. On the other hand, in the Reage program unemployment duration of treated individuals looks to have increased up to 1/5 of a month, although the point estimates are not statistically significant.

Overall, even though there are differences between destination states and programs, the general assessment picture that emerges is that both programs had small impacts on the duration of unemployment spells of treated unemployed. Older workers benefited more

from the program in terms of job placements, whereas younger workers benefited more in terms of training transitions, which were clearly delayed for older workers.

5.2. Heterogeneous effects

We will consider three levels of heterogeneity that might result in differentiated impacts: (i) gender, (ii) age, and (iii) education levels. We will report only DDM estimates based on the kernel matching procedure described earlier. This procedure is applied to each of the subsamples, resulting effectively in exact matching on each of those individual characteristics. Results regarding training are omitted due to the small sample sizes for some of the subsamples.

5.2.1. By gender

We start by exploring a traditional dimension of heterogeneous effects, namely, gender differences. The results suggest that the programs had quite differentiated effects on both genders. However, the results do not indicate that one gender group always benefited more than the other; on the contrary, the sign of the effects alternate depending on the triplet gender, destination state and program. The first two columns of Table 3 report the results.

We have shown earlier (Table 2) that older workers' unemployment spells were reduced before entering a job, while younger workers were affected negatively by the treatment. The current evidence shows that the sign of the impacts was totally driven by the performance of the male population. Men exposed to Reage benefited (-0.85), while those in Inserjovem had spells lasting approximately one more month. Women's performance on the job market as measured by the length of their unemployment spells was not enhanced at all by either program.

In terms of the cancelled registrations, we observed later cancellations in Reage and earlier ones in Inserjovem. The gender results indicate that such results were, respectively, due to the later cancellations for older men and the earlier ones observed for younger women.

5.2.2. By age

The range of ages targeted by the Reage program is rather wide, from 25 to 64 years old. The impact of the program could, therefore, have been different according to the age distribution. To explore such differences, we split our Reage sample into three subsamples: (i) individuals aged less than 30; (ii) individuals aged 30 to less than 40, and (iii) older individuals. The results presented in columns (3) through (5) of Table 3 show some degree of heterogeneity.

The results in terms of 'placed' individuals suggest that the reduction in unemployment durations observed overall (see Table 2) was driven by the good performance of the two younger groups, particularly the middle group. For the older group, 40 or more years old, we did not observe a change in their durations attributable to the program. Regarding exits due to cancellations, the group dominating the overall results is now the oldest one. These individuals, perhaps due to the higher penalizations that cancelling a registration entails for them, were the ones that saw their durations in unemployment increase by 0.83 months due to their participation in Reage. The system was, however, more alert with regards to the younger groups, and this may have resulted in shorter spells, about -0.9 to -0.6 months.

5.2.3. By education level

Finally, in the spirit of the heterogeneous effects identified for the age groups, we now explore the existence of heterogeneous effects by education level. As already reflected in Table 1, the Portuguese labor market is characterized by the predominance of low educational levels, particularly among older individuals (Reage). Thus, given this accentuated duality between educational 'haves' and 'have nots', it is

¹⁰ In the D-in-D framework, following a referee's suggestion, we have also experimented with interacting the treatment dummy variable with the covariates to account for heterogeneous impacts. The results do not differ substantially from the ones reported here. Indeed, computed at the average value of the characteristics, the impact for the Reage program is -0.23 months and for the Inserjovem program 0.07 months; the DDM estimates were -0.40 and 0.29 , respectively (Table 2). The full set of estimates is available from the authors upon request.

Table 3

Short-term average treatment effect on the treated (in months): heterogeneity analysis by gender, age and education levels

Exit type and program	Gender		Reage cohorts (age in years)			Education levels				
	Women	Men	[25, 30]	[30, 40]	≥ 40	4 or less	6 years	9 years	High school	College
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Placed										
Reage	0.02 (0.63)	−0.85 (0.70)	−0.19 (1.14)	−0.49 (0.76)	0.03 (0.76)	0.08 (0.71)	−0.45 (1.00)	−1.95 (1.34)	−1.07 (1.44)	−2.77 (2.35)
	9543	8780	2789	6701	8212	9082	3816	2485	2067	826
Inserjovem	−0.01 (0.49)	0.98 (0.59)	–	–	–	0.11 (1.54)	0.23 (0.73)	0.25 (0.75)	0.66 (0.70)	0.24 (1.17)
	7318	4626	–	–	–	998	3682	2732	3519	1001
Cancelled										
Reage	−0.04 (0.34)	0.45 (0.35)	−0.88 (0.53)	−0.61 (0.42)	0.83 (0.39)	0.84 (0.37)	−1.07 (0.57)	−0.45 (0.74)	−0.72 (0.81)	0.04 (1.23)
	37,643	31,317	12,668	22,325	32,240	36,754	13,670	8407	6825	3304
Inserjovem	−0.59 (0.24)	0.07 (0.30)	–	–	–	−0.59 (0.66)	0.07 (0.35)	−0.26 (0.34)	−0.45 (0.30)	−1.10 (0.50)
	34,400	21,195	–	–	–	5135	15,749	12,601	17,015	5094

Note: Estimates based on the D-in-D matching estimator with kernel propensity score matching. Analytical standard errors reported in parentheses. Total number of observations (treatment and control, before and after) reported below standard errors.

plausible that the programs had differentiated impacts for these two broad groups. The final 5 columns of Table 3 repeat the exercise, but now individually for each of the 5 education levels.

With regards to the 'placed' in Inserjovem, there are no significant differences across the educational levels and the point estimates are close to the average reported in Table 2, an average treatment effect of 1/4 of a month longer spells. However, for the more heterogeneous group in Reage, there are substantial differences. It is evident that the group of the better educated treatment units was the ones benefiting the most from the program, while those with 6 or less years of education either did not benefit at all or experienced tiny decreases in the unemployment spells. Although statistically none of the estimates is significant, we observe reductions in unemployment spells for those with 9 or more years of education ranging from 1 month to almost 3 months.

Finally, there is no pattern identifiable by education level with regards to the treatment effect on the treated for exits due to cancellations.

5.3. Timing effects

The effects of public policies might be anticipated by individuals or delayed in time by a process of learning or adjustment to the new policy. To assess the importance of such effects, we will first study the impact of the program at later periods, namely, in the second semester after implementation and two years after and then study the possibility of anticipation effects.

5.3.1. Impact on the 2nd semester and after two years of implementation

The implementation of the two programs involved putting into place a large and heterogeneous set of resources. It is possible that the coordination and fine tuning of the program took longer than one semester of implementation. This suggests the possibility that an evaluation of the impact of the program at later periods might produce different assessments. For that purpose, we re-evaluate the programs during the 2nd semester of implementation (January through June, 1999)¹¹ and after 2 years (July through December, 2000).¹² The results are presented in the 2nd and 3rd columns of Table 4, respectively.

¹¹ The treatment group now includes individuals from the same pilot EOs, but who received treatment during this later period (2nd semester of implementation); the control group includes individuals from EOs that were not offering the programs during the same semester.

¹² The same logic of construction is used here but for a later period, i.e., from the baseline control group we excluded all EO's that enrolled in the program from December 1998 to June 2000.

While, with the exception of training, the results for the 2nd semester of implementation are similar to the ones for the 1st semester (baseline), the results after two years indicate, without exception, that, on average, participants in both programs spent longer periods unemployed. In both re-evaluations, the statistical imprecision of the estimated has also increased, rendering non-significant the point estimates. Furthermore, the latter evaluation must be taken with some additional caveats. During the intervening two years other policy changes came into effect that might have contaminated the evaluation; perhaps the most prominent, in July 1999, when the entitlement periods of different unemployment insurance benefits were extended for the vast majority of the unemployed (see Centeno and Novo, 2007 for a detailed discussion of this reform and an evaluation of its impact on the duration of subsidized unemployment spells). This entitlement extension might have crowded out the positive effects that ALMP intended to have on the duration of unemployment spells.

5.3.2. Anticipation effects

Anticipation effects may occur within the pilot and/or non-pilot areas. Individuals in the pilot area might have adapted their job-search behavior in anticipation of future participation and the same might have occurred with non-pilot area individuals prior to their EOs enrollment in the programs. In our setting, the first type of anticipation effects are less likely to have occurred because there were no pre-announcements of the specific roll out of the program throughout the country. We also believe that anticipation effects are unlikely to have occurred since in the non-pilot area the next wave of EOs to join took place only in February 1999, 8 months after the pilot EOs.

Nonetheless, to address concerns with the possibility of anticipation effects in the control group, we exclude from this group individuals belonging to the EOs that joined the programs in February and May 1999. The re-estimated impacts are presented in the last column of Table 4. The Inserjovem outcomes are within the range of the estimates presented previously. With regards to the Reage participants, the only result worth highlighting is the one for individuals placed in a job; the new estimate yields a reduction in unemployment duration of 1.42 months. Relatively to the baseline estimate (−0.4, column (1)), this difference is attributable to individuals excluded from this evaluation, who had average unemployment duration smaller in about one month than those in the EOs remaining in the new restricted control group. This reveals that individuals in the baseline control group might have indeed intensified their job-search efforts in anticipation of participation in the new programs.

Table 4

Average treatment effect on the treated (in months): timing effects

Type of exit	Program	1st semester of implementation ¹	2nd semester of implementation ²	Medium term evaluation ³	Anticipation effects ⁴
		(1)	(2)	(3)	(4)
Placed	Reage	–0.40 (0.44) 18,323	–0.26 (0.84) 14,068	1.11 (0.93) 10,875	–1.42 (0.89) 11,703
		0.29 (0.36) 11,944	–0.39 (0.74) 9287	0.96 (0.82) 6586	0.33 (0.80) 7213
		1.70 (1.59) 2182	–0.04 (1.60) 1925	1.23 (1.77) 1.644	2.49 (1.92) 1445
Training	Reage	–0.21 (0.78) 3168	0.41 (1.07) 2620	1.27 (1.09) 2098	0.08 (1.18) 2106
		0.13 (0.28) 68,960	0.53 (0.62) 53,078	0.41 (0.70) 38,796	0.13 (0.66) 44,598
		–0.26 (0.18) 55,595	–0.58 (0.51) 42,142	0.04 (0.57) 27,692	–0.63 (0.53) 35,809

Notes: Bootstrapped-based standard errors are presented in parentheses, followed by the number of observations. (1) This column presents the D-in-D matching estimates with kernel propensity score matching reported in column (3) of Table 2. (2) The impact of the program is measured during the second semester of implementation. (3) The impact of the program is measured during July to December 2000, 2 years after the first EOs started offering the programs, and before the programs are rolled out to the whole country (the last EOs enrolled in January, 2001), using the same set of treatment EOs and dropping from the control group EOs that enrolled between December 1998 and July 2000. (4) The impact of the program is measured during the first semester of implementation but, relatively to the baseline specification, the control group excludes EOs that were in line to offer the programs, namely, those joining in February and May 1999.

5.4. Local labor markets: substitution effects and comparability issues

We argued that the control units came from EOs with similar labor market characteristics (see Fig. 3) and that there was no premeditated selection behavior on the part of the program administrators (see Figs. 1 and 2). Among others, Heckman et al. (1997, 1998) have stressed the importance of using treated and control units from the same local labor market to conduct causal evaluation. To further address this concern, we consider other constructions of our treatment and control groups.

5.4.1. Treatment units: excluding Alentejo

In Fig. 1, there is one noticeable feature that may hinder the local comparability of treatment and control groups: in the southern region of Portugal, called Alentejo, there are only treatment EOs.¹³ This region has specific characteristics that may be harder to replicate elsewhere. In particular, Alentejo is characterized by slightly higher unemployment rates, and by large proportions of employment in the public and agricultural sectors, although the region is not densely populated.

Thus, we redefine our treatment group by excluding individuals who received treatment in the Alentejo EOs. The results of this evaluation are presented in column (3) of Table 5, alongside column (1), which shows the already reported DDM estimates (kernel matching) of Table 2.¹⁴

The exclusion of Alentejo could influence the estimates of the program impact, particularly in terms of transitions to employment. Notice, however, the remarkable similarity to the ‘placed’ results discussed hitherto (first column). In the two other instances the impacts are larger – cancelled registration in Inserjovem (an effect that exceeds the previous by about two weeks) and transitions to training in Reage (unemployment spells lasting 2.7 months longer,

rather than 1.7 months). Notice, however, that the confidence intervals derived under the two constructions overlap and, therefore, statistically we cannot exclude the hypothesis that both estimates are the same. Furthermore, these changes occur in the categories that are less subject to labor market conditions, but are more dependent on the particular conditions, characteristics and discretion of the EOs (e.g. availability of training and how closely the cancellations rules are followed).

Table 5

Short-term average treatment effect on the treated (in months): substitution effects and issues of local comparability

Exit	Program	D-in-D matching	Robustness check by adjusting	
		Kernel (Table 2) ¹	Treatment group ²	Control group ³
		(1)	(2)	(3)
Placed	Reage	–0.40 (0.44) 18,323	–0.39 (0.52) 16,981	–1.01 (0.53) 16,567
		0.29 (0.36) 11,944	0.24 (0.40) 11,003	0.50 (0.41) 10,420
		1.70 (1.59) 2182	2.71 (1.53) 1762	0.06 (2.03) 1651
Training	Reage	–0.21 (0.78) 3168	–0.04 (0.76) 2656	0.15 (0.81) 2522
		0.13 (0.28) 68,960	0.34 (0.28) 65,770	–0.10 (0.29) 60,641
		–0.26 (0.18) 55,595	–0.71 (0.18) 52,822	–0.27 (0.19) 47,492

Notes: Bootstrapped-based standard errors are presented in parentheses, followed by the number of observations. (1) This column presents the D-in-D matching estimates with kernel propensity score matching already reported in column (3) of Table 2. (2) The robustness check is performed by adjusting the definition of the treatment group. All EOs in the southern region Alentejo are excluded from the treatment group due to local labor market specificity (higher unemployment rates and higher portions of agricultural and public service employment). (3) The robustness check is performed by considering a stricter definition of eligible individuals for the control group, specifically, only those residing in EOs having physical borders with the EOS in the treatment group.

¹³ A few EOs, located in Alentejo, joined in October 1998, only 4 months after the pilot EOs and within our 6 months evaluation time window, resulting in the exclusion of these EOs from our control group.

¹⁴ We replicate the estimation procedures underlying Table 2 for this new construction. Although omitted from Table 5, the choice of the estimation methodology, as in Table 2, does not drive our results. As before, we only report the DDM estimates (kernel based propensity score matching).

5.4.2. Control units: a subset of EOs

We now consider as control units only those individuals drawn from non-pilot EOs areas with common physical borders with the pilot EOs area. This may have three advantages over the previous construction. First, EOs come from similar local labor markets, specifically in terms of the average unemployment rate, without any of the pitfalls that may arise when comparing mean values constructed with more diverse labor markets. Secondly, more than 98% of this control sample comes from EOs that joined the program on or after May 1999 (almost a year after the pilot EOs). Therefore, it is rather unlikely that individuals and/or EOs adapted their behavior in anticipation of the programs. Finally, and more important, given the geographical and economic proximity of the EOs, we can gauge the importance of substitution effects. In fact, if substitution effects are important, we expect the results based on this counterfactual group to be larger than the previous ones, reflecting the spillover in the labor market in favor of treated individuals (who would have shorter unemployment spells compared to those of the more substitutable workers in neighboring labor markets).

The results presented in the last column of Table 5 have remarkable qualitative and quantitative similarities with the initial results (and the above robustness check). The only quantitative exception is the impact on Reage individuals placed in a job. For this transition category, we now have point estimates that can reach up to minus one month of unemployment. This result may point towards a small substitution effect. Interestingly, however, the substitution effect on youth points, if anything, towards longer unemployment spells when considering neighbor EOs.

In view of the underlying distribution of unemployment durations, which has rather long spells, the set of results herein presented, which include several layers of heterogeneity and robustness checks, seems to lead to the conclusion that the impacts on the unemployment durations of the treated were rather paltry.

6. Conclusions

This paper evaluates the short-run impact of Inserjovem and Reage, two ALMP initiatives, introduced in Portugal in the late 1990s. The measure of program effectiveness is the reduction in unemployment duration through program participation. The programs target two distinct age groups: the Inserjovem, for individuals under 25; the Reage, for individuals 25 or over. These programs were mandatory with treatment ranging from job-search assistance to training and vocational or non-vocational guidance. No wage subsidies were included in the possible treatments. The possibility of analyzing the response of young and older workers to such a treatment in the same labor market is one of the main advantages of the Portuguese experience.

Identification of the average treatment effect on the treated is achieved by means of an important source of variation generated by the area-based pilot implementation of the programs across the country. This implementation created natural treatment and control groups, drawn from pilot and non-pilot areas, respectively.

The positive impact of the programs in reducing unemployment duration is very limited. In fact, through program participation, we find a small reduction on the length of unemployment spells for Reage participants finding a job, and a tiny increase in the spell durations of individuals in Inserjovem. These results are driven by the behavior of men, as women seem to react less to these treatments. Our results point to a more positive impact for individuals aged 30 to 40 and among the better educated.

An important lesson to be drawn from this study for the ongoing discussion on ALMP for older and disadvantaged workers is the apparent difficulty of the programs in improving their unemployment experiences. Indeed, the less educated individuals and those over 40 did not benefit at all from the programs, which can be seen as a partial failure of the programs with two of the groups that face the worst

prospects in modern labor markets. The impact of the program at later points in time, admittedly after a learning period by EOs and other stakeholders, did not improve the prospects of the unemployed, either.

As with job placement, the programs also had a differentiated impact on the other two destination states considered. In particular, the programs reduced unemployment duration for young workers exiting the labor force, but increased it for older workers. A similar result is obtained with exits to training.

What can explain the weakness of these impacts when compared with other recently reported in the literature (e.g. Blundell et al., 2004; Di Giorgi, 2005 for the United Kingdom)? A possible driving force can be traced back to the work of Katz (1998), who, in reviewing different ALMP, found that the most successful were those combining wage subsidies with job-search assistance and training. Contrary to the policies implemented, for instance, in the United Kingdom, the first dimension was absent in the Portuguese policy mix. Katz (1998) also showed that such a combination was particularly helpful for more disadvantaged groups, exactly those who did not benefit from the Reage and Inserjovem programs.

We looked only at one dimension of the programs' success. There are, however, alternative dimensions for future research paths, namely longer-run effects and post-unemployment job match quality.

Acknowledgements

We are grateful for the comments from the editor and two anonymous referees. We also thank participants at the 2004 LoWER Conference, London, the 2004 EALE Annual Conference, Lisbon, the 2004 AIEL Conference, Modena and at the HIVA Conference, Brussels. We acknowledge the collaboration given by the Portuguese employment agency, IEFP, in particular, for access to and detailed discussion of the data. The financial support of FCT (Fundação para a Ciência e a Tecnologia), Portugal, through UECE (Research Unit on Complexity and Economics) is acknowledged. This article is part of the Multi-annual Funding Project (POCTI/0436/2003). The opinions expressed herein are not necessarily those of the Banco de Portugal.

Appendix A

A.1. The matching procedure – common support

The implementation of a DDM estimator for repeated cross-section data entails imposing several layers of common support to guarantee that treatment and control units are comparable (have common support) in the pre- and post-program periods. To achieve these, we proceeded as follows:

1. The matching of treatment and control units in the post-program period is achieved after imposing a common support condition, i.e., that the range of estimated propensity scores is the same for treated and control units;
2. Then, we match treatment and control units in the before period, imposing still a common support condition between these two groups;
3. The successfully matched units in steps 1) and 2) are further restricted to have propensity scores in a common range. This guarantees that all units have common support of propensity scores.

In steps 1) and 2) each treatment unit's estimated propensity score is matched to the control unit(s) propensity score(s), according to the matching method of choice – kernel or spline-smoothing.

An alternative scheme is implemented in Blundell et al. (2004).

In practice, the propensity scores estimated for the 4 groups (Treatment/Control, Before/After) ended up being rather similar. This is not strange to the fact that individuals in the treatment and control group, even before matching, shared rather similar average pre-

Table A1

Propensity scores: estimation based on probit models for placed individuals by program and period

Variables	Reage				Inserjovem			
	After		Before		After		Before	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Age	−0.05	(0.01)	−0.02	(0.01)	0.22	(0.14)	0.24	(0.14)
Age ²	0.00	(0.00)	0.00	(0.00)	−0.01	(0.00)	−0.01	(0.00)
Married	0.10	(0.04)	0.10	(0.04)	0.08	(0.05)	0.12	(0.05)
Female	0.01	(0.03)	0.10	(0.03)	−0.02	(0.04)	0.01	(0.04)
Education ⁽¹⁾								
4 or less years	0.79	(0.09)	0.67	(0.08)	0.60	(0.10)	0.29	(0.10)
6 years	0.57	(0.09)	0.55	(0.08)	0.49	(0.09)	0.27	(0.08)
9 years	0.38	(0.09)	0.40	(0.09)	0.49	(0.09)	0.27	(0.08)
High school	0.35	(0.09)	0.33	(0.09)	0.41	(0.08)	0.20	(0.08)
Reason to register ⁽²⁾								
Inactive	0.27	(0.10)	0.26	(0.12)	0.26	(0.05)	0.25	(0.05)
Quitted	−0.14	(0.07)	−0.11	(0.08)	0.01	(0.10)	−0.08	(0.10)
Others	0.09	(0.04)	0.15	(0.03)	0.23	(0.06)	0.14	(0.06)
Immigrant	−0.35	(0.17)	−0.38	(0.18)	0.14	(0.24)	−0.33	(0.27)
Handicapped	0.38	(0.20)	−0.18	(0.18)	−0.03	(0.25)	0.34	(0.25)
Region ⁽³⁾								
North	−1.29	(0.04)	−1.15	(0.04)	−1.24	(0.04)	−1.24	(0.04)
Center	−0.01	(0.05)	−0.04	(0.04)	−0.17	(0.05)	−0.27	(0.05)
Algarve	−0.87	(0.11)	−1.02	(0.09)	−0.96	(0.12)	−1.14	(0.10)
Constant	0.24	(0.31)	−0.55	(0.29)	−2.81	(1.54)	−2.87	(1.53)
No. observations	8469		9854		5724		6220	
Pseudo R ²	0.15		0.13		0.14		0.14	

Notes: See Table 1 for variable definitions. (1) The omitted categorical variable is “College”. (2) The omitted categorical variable is “Fired”. (3) The omitted categorical variable is “Greater Lisbon”. The dummy variable for Alentejo was also omitted from the estimation because it predicted perfectly treatment (it should be noted that the vast majority of Alentejo EOs applied the programs in June 1998, while none of the exceptions qualified for the control group).

treatment characteristics (see Table A1); for the units in the before period, a similar exercise showed a similar pattern across pre-treatment characteristics. Furthermore, the common support exercise benefits from a large pool of individuals from which to draw units with ‘common’ characteristics. Therefore, the algorithm assured common support across all four group, dropping from the total sample only a few observations (Table A2 at the bottom reports common support information for the Reage group of placed individuals in the post-treatment period).

A.2. Estimating the propensity score

One of the cornerstones of a successful matching procedure is the quality of the estimated propensity scores. In turn, the estimation of the propensity scores entails two choices: (i) the type of model and (ii) the choice of the variables to include in the model specification. While the choice of the model type is rather standard, revolving around the ubiquitous probit or logit models, the choice of the set of variables to be included is less consensual and more crucial to a successful use of the propensity score matching models. While some studies, Augurzy and Schmidt (2001) and Bryson et al. (2002), advise against overparameterized models, Rubin and Thomas (1996) recommend against parsimony.

For the first part, we settled for the estimation of probit models. For the set of covariates reported in Table A.2, which include only variables that influence simultaneously the outcome and the treatment status (Smith and Todd, 2005; Sianesi, 2004), we relied on economic theory, previous research and the institutional setting. Extensive literature in labor economics has demonstrated that characteristics such as education level, age and previous labor market status (active/inactive, quitted/fired) are important determinants in the transitions out of unemployment. Thus, the inclusion of such variables in our model is justified because it certainly affects the outcome variable, but also because it is rather likely that if there was some type of selection on observables (on the part of administrators), then it was also based on such characteristics. Our model specifications further include variables such as marital status, gender, disability and immigrant dummies, which do affect labor market transitions. Finally, we controlled for regional differences, which may be less important in

determining the treatment status, but are clearly important to control for differences in the performance of regional labor markets. All the covariates are measured before the treatment is administered.

A.3. Assessing the matching quality

We assess the balancing of the distribution of covariates used in the propensity scores estimation within each of the matching exercises. Table A2 reports the *p*-value of the standard *t*-test for the equality of mean sample values, the standardized bias (Rosenbaum and Rubin, 1985), and the joint significance tests and pseudo-*R*² (Sianesi, 2004). This table details only the assessment of the sample of placed individuals in the Reage program in the post-program period.¹⁵

Starting with the standard *t*-tests for the equality of means for the treatment and control groups for each of the variables included in the specification of the propensity score model, there are two noteworthy results. First, as already pointed out in Table 1, there is a remarkable similarity between the treatment and control group's mean values before matching. Nonetheless, from a pure statistically point of view, the mean values are different from each other for the majority of variables. This leads to the second point. After matching, we fail to reject the null hypothesis of mean equality between the treatment and control groups for all variables.

In the next columns, we report the standardized bias as proposed by Rosenbaum and Rubin (1985), and the reduction in the absolute bias obtained after matching control and treatment units. One pitfall of this indicator often pointed out is the fact that it has no formal (statistical) threshold for assessing the success of the reduction in mean bias. Nonetheless, the values that we obtain seem to grant some success in the matching procedure, with reductions as large as 98.7%, and are in line with previous empirical studies (Lechner, 1999; Sianesi 2004). At the bottom of the table, we also report some overall summary statistics for the absolute bias reported. The mean absolute bias in the matched sample is 1.82, while in the unmatched sample is 14.64.

¹⁵ The remaining results are available from the authors upon request; qualitatively they are similar to the ones discussed herein.

Table A2

Matching covariates balancing property: summary statistics and test statistics for placed individuals by program in the post-program period

Variables	Reage		<i>t</i> -test <i>p</i> -value ⁽¹⁾	% Bias ⁽²⁾	Reduction in bias
	Sample	Mean			
		Treated			
Age	Unmatched	40.67	39.71	0.000	
	Matched	40.68	40.96	0.317	−2.9 71.3
Age ²	Unmatched	1745.80	1661.80	0.000	
	Matched	1747.10	1767.80	0.374	−2.6 75.4
Married	Unmatched	0.75	0.73	0.030	
	Matched	0.75	0.76	0.834	−0.6 88.7
Female	Unmatched	0.56	0.53	0.003	
	Matched	0.56	0.55	0.221	3.5 51.3
Education					
4 or less years	Unmatched	0.56	0.47	0.000	
	Matched	0.56	0.56	0.966	−0.1 99.4
6 years	Unmatched	0.20	0.21	0.124	
	Matched	0.20	0.20	0.732	−1 73.8
9 years	Unmatched	0.12	0.14	0.001	
	Matched	0.12	0.12	0.777	0.8 90.4
High school	Unmatched	0.10	0.12	0.001	
	Matched	0.10	0.10	0.851	−0.5 93.9
Reason to register					
Inactive	Unmatched	0.03	0.02	0.066	
	Matched	0.03	0.02	0.255	3.3 23.3
Quitted	Unmatched	0.05	0.07	0.000	
	Matched	0.05	0.05	0.360	−2.4 75.5
Others	Unmatched	0.28	0.23	0.000	
	Matched	0.28	0.26	0.205	3.7 68.6
Immigrant	Unmatched	0.01	0.01	0.161	
	Matched	0.01	0.01	0.976	0.1 97.8
Handicapped	Unmatched	0.01	0.00	0.114	
	Matched	0.01	0.01	0.747	0.9 74.7
Region					
North	Unmatched	0.10	0.48	0.000	
	Matched	0.10	0.11	0.576	−1.2 98.7
Center	Unmatched	0.16	0.08	0.000	
	Matched	0.16	0.17	0.325	−3.2 85.8
Algarve	Unmatched	0.01	0.03	0.000	
	Matched	0.01	0.01	0.267	2.3 82.8
Observations					
On common support		2421	6045		
Off common support		3	0		
	Unmatched		Matched		
Bias summary statistics					
Mean		14.64	1.82		
S.D.		21.39	1.30		
Maximum		92.15	3.73		
Minimum		3.51	0.08		
Pseudo R ²⁽³⁾		0.150	0.001		
LR test <i>p</i> -value		0.000	0.865		

Notes: (1) The p-value of the t-test for the equality of means in the treated and control groups, both before and after matching. (2) % Bias is the standardized bias as suggested by Rosenbaum and Rubin (1985) reported together with the achieved percentage reduction in |bias|. (3) Pseudo R² from the probit model estimation of the propensity scores, including all variables reported in Table A.1, before and after the matching process (Sianesi, 2004).

Finally, following the suggestion of Sianesi (2004), we look at the joint statistical significance of the covariates and at the pseudo-R² of the propensity score in the unmatched and matched samples estimation procedures. As it can be seen in the last two rows of Table A.2, the pseudo-R² in the propensity score estimation that used only the treated units and the matched control units falls to values close to zero. The F-test complements this information, corroborating the view that matching has successfully eliminated any systematic observable differences between the treated and control groups.

References

- Abadie, A., 2005. Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72 (1), 1–19.
- Addison, J., Centeno, M., Portugal, P., 2004. Reservation wages, search duration, and accepted wages in Europe. Working paper 13-04. Banco de Portugal.
- Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. *Review of Economic and Statistics* 60, 47–57.
- Agurzkzy, B., Schmidt, C., 2001. The Propensity Score: A Means to an End. IZA, Working paper.
- Becker, S.O., Ichino, A., 2002. Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2 (4), 358–377.
- Blundell, R., Costa Dias, M., 2000. Evaluation methods for non-experimental data. *Fiscal Studies* 21 (4), 427–468.
- 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* 2 (4), 569–606.
- Bonnal, L., Fougère, D., Sérandon, A., 1997. Evaluating the impact of French employment policies on individual labour market histories. *Review of Economic Studies* 64, 683–713.
- Bryson, A., Dorsett, R., Purdon, S., 2002. The use of propensity score matching in the evaluation of active labour market policies. Working Paper 4, Policy Studies Institute and National Centre for Social Research.
- Caliendo, M., 2006. Microeconomic evaluation of labour market policies. Lecture Notes in Economics and Mathematical Systems. Berlin Heidelberg, Springer-Verlag.
- Carling, K., Richardson, K., 2004. The relative efficiency of labor market programs: Swedish experience from the 1990s. *Labour Economics* 11 (3), 335–354.
- Centeno, M., Novo, Á.A., 2007. Identifying unemployment insurance income effects with a quasi-natural experiment. Working paper 10/2007. Banco de Portugal.
- De Giorgi, G., 2005. The New Deal for young people five years on. *Fiscal Studies* 4157 (3), 371–388.
- Dehejia, R., Wahba, S., 1999. Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* 94 (448), 1053–1062.
- Dehejia, R., Wahba, S., 2002. Propensity score matching methods for non-experimental causal studies. *The Review of Economics and Statistics* 84 (1), 151–161.
- Eberwein, C., Ham, J., LaLonde, R., 2002. Alternative methods of estimating program effects in event history models. *Labour Economics* 9 (2), 249–278.

- Eichler, M., Lechner, M., 2002. An evaluation of public employment programmes in the East German state of Sachsen-Anhalt. *Labour Economics* 9 (2), 143–186.
- Gerfin, M., Lechner, M., 2002. A microeconomic evaluation of the active labour market policy in Switzerland. *The Economic Journal* 112 (482), 854–893.
- Heckman, J., Hotz, J., 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *Journal of the American Statistical Association* 84 (408), 862–874.
- Heckman, J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions: an overview. *Journal of Econometrics* 30, 239–267.
- Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* 64 (4), 605–654.
- Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. *Econometrica* 66 (5), 1017–1098.
- Heckman, J., LaLonde, R., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 3A. North-Holland, Amsterdam, pp. 1865–2097.
- Katz, L., 1998. Wage subsidies for the disadvantaged. In: Freeman, R., Gottschalk, P. (Eds.), *Generating jobs: How to increase demand for less-skilled workers*. Russell Sage Foundation, New York, pp. 21–53.
- Kluve, J., 2006. The effectiveness of European active labor market policy. Working paper 2018, IZA.
- LaLonde, R., 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76 (4), 604–620.
- Lancaster, T., 1992. *The Econometric Analysis of Transition Data*. Cambridge, Cambridge University Press.
- Larsson, L., 2003. Evaluation of Swedish youth labor market programs. *The Journal of Human Resources* 38 (4), 891–927.
- Lechner, M., 1999. Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business & Economic Statistics* 17 (1), 74–90.
- Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. URL <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Meyer, B.D., 1995. Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* 13, 151–162.
- Michalopoulos, C., Bloom, H., Hill, C., 2004. Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *The Review of Economics and Statistics* 86 (1), 156–179.
- OECD, 2006. *Live Longer, Work Longer*. Paris, OECD.
- Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.
- Rosenbaum, P., Rubin, D., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39 (1), 33–38.
- Rubin, D., 1977. Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics* 2, 1–26.
- Rubin, D., 2006. *Matched Sampling for Causal Effects*. Cambridge University Press.
- Rubin, D., Thomas, N., 1996. Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52 (1), 249–264.
- 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., Todd, P., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125 (1–2), 305–353.
- van den Berg, G., Klaauw, B.v.d., 2006. Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment. *International Economic Review* 47 (3), 895–936.