Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Emil Mihaylov

Evaluation of Subsidized Employment Programs for Long-Term Unemployment in Bulgaria

A Matching Approach

MSc Thesis 2009

Evaluation of Subsidized Employment Programs for Long-Term Unemployed in Bulgaria: A Matching Approach

Master thesis

Thesis:	MSc Economics, July 2009
Student:	Emil Mihaylov
Student number:	1530135
Email address:	e.s.mihaylov@student.vu.nl
Supervisor:	Prof. Dr. Bas van der Klaauw
Second reader:	Prof. Dr. Annette Bergemann
Faculty:	Faculty of Economics and Business Administration

vrije Universiteit amsterdam

Acknowledgements

At first place, I would like to thank my supervisor Bas van der Klaauw for his time, comments and excellent supervision. At second place, I would like to thank the second reader of my thesis, Annette Bergemann. I gratefully acknowledge her assistance in finding a suitable topic for my thesis. Bas and Annette, thank you very much both!

I would like to thank also Barbara Sianesi for helping me find my way in PSMATCH2. My special thanks go to my family for their support. Last but not least, I would like to thank God for being able to finish this thesis.

Emil Mihaylov

Table of contents

Chapter 1: Introduction

Bulgaria's transition to a market economy has been remarkably slow and painful. Difficulties that can affect the labor markets of transition economies have all occurred in Bulgaria. Sharp declines in employment, high unemployment, low turnover among the unemployed and increasing long-term unemployment are characterizing for the Bulgarian labor market after 1989 (OECD, 1998). Following the closure of state enterprises, total employment fell dramatically in the first decade after 1989. From nearly 4 million in 1989 total employment decreased to slightly over 3 million in 1995 (Garibaldi et al., 2001). From virtually absent under central planning, unemployment rose to over 16 percent in 1993 and nearly 20 percent in 2001.

A striking feature of the transition process to a market economy in Bulgaria is the long average duration of unemployment (OECD, 1998). Low outflow rates from unemployment resulted in unemployment persistence and large shares of long-term unemployment. In 1993 around 50 percent of the unemployed individuals were continuously unemployed for more than one year. In 1996 this percentage was 75, which decreased to around 60 percent in 2000 and it has been stable since then. Individuals with unfavourable demographic and skill characteristics, i.e. old, ethnic or low qualified, are overrepresented in the pool of long-term unemployed.

In order to reduce averse effects of unemployment, the Bulgarian government introduced a broad array of active labor market programs (ALMPs) in the early 1990s (De Koning et al., 2007). Despite limited resources, ALMPs have been used on a large scale in Bulgaria, including most programs that are in place in developed countries. Measures as labor market training, subsidized employment, public works, programs for youth and disable constitute the most important among the labor market programs.

This paper makes a modest attempt to evaluate the effectiveness of employment subsidies for long-term unemployed. In particular, the aim of the paper is to answer the following research question:

Are subsidized employment programs an effective instrument to bring long-term unemployed workers back to work?

Main contribution of this paper is that it enriches the scarce empirical evidence on the effectiveness of active labor market policies in Bulgaria. Previously only two studies have made an effort to perform such evaluations. De Koning et al. (2007) examine the impact of a large-scale public work program, called "From social assistance to employment" and Walsh et al.(2001) evaluate various youth employment programs.

Using a relatively rich (for an Eastern European country) administrative and survey dataset, this paper assesses the effects of the program on the employment chances of the participants. A matching estimator, based on propensity scores is used to estimate treatment effects by comparing employment outcomes of participants and non-participants. After adjusting for observable differences between participants and non-participants, a positive and highly significant treatment effect of the program is estimated. The employment rate of the participants is about 33 percentage points higher than that of the matched control group. Additional analysis reveals that the program is effective also for various subgroups in the sample. The estimated result is not sensitive to different definitions of the outcome variable and different matching estimators. Further, the estimated program effect was scrutinized to see if it is sensitive to hidden bias, arising from unobserved factors affecting simultaneously assignment to treatment and the outcome variable. The sensitivity analysis was carried out in the framework of the Rosenbaum bounds approach. The analysis reveals that the estimated treatment effect is also not sensitive to hidden bias.

The rest of the paper is organized as follows. Chapter 2 describes shortly the institutional framework in Bulgaria. Chapter 3 gives an overview of previous evaluation studies. Chapter 4 sets the framework for the empirical analysis. It starts with a discussion of the main evaluation problem – each individual has two potential outcomes (treatment and non-treatment), whereas only one is observed. A challenge for each non-experimental study is to find an approximation of the counterfactual outcome. This chapter discusses also how the evaluation problem is solved in this study: by assuming conditional mean independence and applying propensity score matching techniques. Chapter 5 provides a description of the dataset and variables used in this analysis. Most importantly, this chapter discusses the plausibility of the conditional mean independence assumption on the hand of the available data. Chapter 6 discusses the implementation of the matching estimator. The chapter provides information on the estimated propensity score and also some matching details. Chapter 7 presents the main results. Chapter 8 discusses the importance of selection bias due to unobservables for the estimated treatment effect in chapter 7. The estimate in chapter 7 crucially relies on the conditional mean independence assumption. However, if there is a positive unobserved selection into treatment, meaning that those who are more likely to participate are also more likely to be employed, given the same observables, then the estimated treatment will be overestimated. The purpose of this chapter is to examine how strong unmeasured influences have to be so that the estimated positive treatment effect is purely due to selection effects. Finally, chapter 9 concludes.

Chapter 2: Institutional framework

The first part of this chapter, section 2.1, provides information about the administration and labor market policies in Bulgaria. The second part, section 2.2, describes the subsidized employment program for long-term unemployed.

2.1. Administration and labor market policies in Bulgaria

Administration in Bulgaria

The Ministry of Labor and Social Policy in Bulgaria is the governmental body for development of policies in the field of labor, income and social security. The National Employment Agency (NEA) is the administrative agency responsible for the direct implementation of the labor market policies. The NEA provides organizational and informational support on labor market trends and labor program activities to nine major employment offices, called Regional Employment Service Directorates (RESDs). These RESDs organize and coordinate the implementation of labor market policies at regional level. Through a network of 166 Local Labor Offices (LLOs) job services and program activities are delivered to unemployed workers and employers at local level. Main functions of LLOs are the registration and counseling of unemployed, job brokerage, organization of active labor market programs.

Labor market policies

The labor market policy in Bulgaria is carried out thought administration of both passive and active labor market measures. The main part of the passive measures is unemployment compensation. Bulgaria has a quite tight unemployment benefits system, which results in a relatively small share of passive measures in total spending on labor market policies (see Table 1). To qualify for unemployment benefits individuals have to be employed for at least 9 months during the last 15 months and have been subject to compulsory social insurance coverage. The duration of unemployment benefits is dependent on the duration of the previous employment spell. It varies from 4 months (for unemployed with previous employment spell between 9 months and 3 years) to 12 months (for unemployed with previous employment spell above 25 years). In addition, individuals have to be registered with the local employment office and are willing to accept any proper job or training, that is offered. The cash benefits amount to 60 percent of the previous wage, but they can not be lower than 70 and higher than 130 percent of the minimum wage. Responsible for the payment of unemployment benefits is the National Social Security Institute (NSSI). Due to the long duration of unemployment spells however, only a small fraction of the registered unemployed receive unemployment benefits[1]. If after the expiration of the unemployment benefits, the individual is still unemployed, he/she may receive social benefits. The social benefits are means-tested and related to the minimum income. They are paid for an unlimited period. Responsible organization for the payment of social benefits is the National Social Assistance Agency.

Table 1: Passive and active labor market policies

	2004	2005	2006	2007
PLMP[a]	0.26	0.21	0,18	0.15
ALMP[a]	0.46	0.43	0.38	0.30
Total participants in ALMP	187,249	173,594	148,723	130,345
ALMP participants as % of all registered unemployed	39.9	40.9	41.7	45.4

Source: Eurostat and National Employment Agency Annual reports
[a] Expenditure as percent of GDP.

Unemployed individuals can participate in various active labor market programs. These programs contain measures such as job search assistance, training, employment subsidies and start-up grants, direct job creation, programs for people with disabilities etc. To participate in ALMPs individuals have to be registered as unemployed. If, one month after their registration they are not offered a proper job, they can participate in one of the labor market programs.

[1] According to statistics of NSSI between 2004 and 2007 less than 20 percent of all registered unemployed were subject to unemployment benefits. This percentage increased to 28 percent in 2008.

Around 130,345 individuals participated in ALMPs in 2007, which is approximately 45 percent of all registered unemployed (see Table 1). Most of the programs are targeted at specific groups and to participate in these programs individuals have to meet certain requirements. This study discusses further only the employment subsidies for long-term unemployed, which are of particular interest for the analyses[2].

2.2. Subsidized employment for long-term unemployed

The subsidized employment program comprises incentive measures for private employers to hire unemployed workers. The incentives have the form of direct wage subsidies and cover all social security contributions made by the employer. The wage subsidies are given for a period of 12 months. The employer in turn is required to guarantee employment to the hired unemployed person for a period of 24 months. During the program participants are not expected to actively search for a regular, non-subsidized jobs. Also participants are no longer counted as unemployed. Entitled for the program are individuals who are continuously unemployed for more than 12 months and are registered with the local employment office. If after the end of the program the participants do not continue their work-relation or do not find another job, they can register as unemployed and use again the services of the employment agency. In principle, participation in the program entitles individuals to claim unemployment benefits for the period in which they were employed. However, due to the short duration and low level of unemployment benefits and the long duration of the program, it is not very likely that the program is used to re-new eligibility for unemployment benefits. Finally, the decision to participate in the program is taken by both the unemployed and the case-worker.

Chapter 3: Literature review

This chapter gives an overview of micro-econometric studies that evaluate the effectiveness of subsidized employment programs[3].

---

[2] For a complete overview of all active labor market programs the interested reader is referred to World Bank (2003).

[3] For a survey of macro-econometric studies see Calmfors et al. (2002).

Using data from the 18th Polish labor force survey, Kluve et al. (1999) assess the impact of (among others) wage subsidies on the employment probability of the participants. Their empirical analysis, carried out in the framework of matching, shows that wage subsidies have an overall negative effect on the employment rate of participants. The authors find that the negative effect is even stronger for men. Men who participate in the program have on average 24 percent lower employment and 24 percent higher unemployment rates in the short-run (six months after program end). For women, the effect of the program is not statistically different from zero, both in terms of employment and unemployment rates. Also the medium-term effects of the program, i.e. nine months after program end, are not favourable. The authors conclude that the main reason for the estimated negative treatment effect is that employment programs in Poland are often used to re-qualify for unemployment benefits.

Differently from Kluve et al. (1999), Forslund et al. (2004) find more favourable results. Based on matching and instrumental variable techniques, the authors find that employment subsidies given to private firms in Sweden have an overall positive treatment effect. As a result of the program the unemployment duration of participants is reduced by 8 months as compared to non-participants. The authors find also small negative (lock-in) effects during the first 7 months of the program, which effects turn into positive after this period and remain positive until the end of the observational window of 56 months.

Zhang (2003) reports similar results for Norway. Using administrative data the author examines the effects of (among others) wage subsidies on the transition rate to employment. Zhang finds small negative treatment effects during the program. After six months the treatment effects turn positive and grow steadily thereafter. In the post-treatment period, Zhang estimates an increase of the hazard rate to employment by around 87 percent for participants. These effects are even stronger for women (which is in accordance with the survey of Bergemann and Van den Berg (2006), who find that hiring subsidies impact positively for women and the estimated effects for women always exceed those for men). Zhang (2003) concludes that wage subsidies are an effective instrument to combat unemployment and enhance job opportunities.

Based on longitudinal data from Denmark, also Bolvig et al. (2003) find small lock-in effects during the program and large positive effects in the post-program period. Bolvig's et al. analysis, carried out in a duration model framework, shows that participation in employment programs significantly increases the transition rate from welfare to work. Additionally, the authors model the optimal time of participation. They conclude that for men an early treatment is preferred, whereas for women the treatment should be postpone – there are severe lock-in effects and post-treatment effects are largest in a later stage of the welfare spell.

Gerfin et al. (2005) evaluate the effectiveness of two Swiss subsidized employment programs. Using a rich administrative dataset, the authors show that both employment programs are effective in bringing the unemployed back to work when the participants are long-term unemployed. The programs are found not to be effective for individuals with shorter unemployment spells and individuals that easily can find a job.

None of these studies however, accounts for ex-ante (or threat) effects of active labor market programs. In this respect, Rosholm and Svarer (2004) argue that ignoring the treat effect may lead to an underestimation of the true treatment effect. The reasoning behind their argument is quite intuitive. In a system with ALMPs, unemployed individuals are threat with participation, conditional on remaining unemployed for a certain period. Assuming that ALMPs are like a tax on leisure for the unemployed, this means that the utility of unemployment is lower in a system with ALMPs as compared to a system without. Everything else being equal, unemployed individuals in a system with ALMPs are expected to search harder for jobs and also to decrease their minimum acceptable wages. This in turn will lead to a higher job finding rate, hence a higher exit from unemployment even before an actual treatment takes place.

The quantitative importance of the treat effect is explicitly accounted for in Rosholm and Svarer (2004). In their study the authors evaluate (among others) the impact of private sector employment subsidies on the unemployment duration of participants. Using Danish administrative data, the authors base their empirical analysis on two econometric models. First, the timing-of-events model for identifying treatment effects is used to simultaneously model the transition rate out of unemployment and the transition rate into an ALMP. Second, the dependent hazard rate model is used to estimate two hazard rates - the hazard rate from unemployment into employment and the hazard rate into an ALMP (where the authors explicitly allow the hazard rate into employment to depend on the hazard rate into an ALMP). By combining these two models, the authors estimate both the threat effects of a system with ALMPs and the lock-in and post-treatment effects of private sector employment subsidies. Rosholm and Svarer find that the threat effect reduces the average unemployment duration by three weeks. The private sector employment subsidies are estimated to have small lock-in effects and fairly large positive post-treatment effects. The authors show that not accounting for the threat effect leads to an overestimation of the negative lock-in effects and underestimation of the positive post-treatment effects of private sector employment subsidies. Based on the results the authors conclude that the employment subsidies are an effective active labor market instrument to get individuals out of unemployment.

Finally, to conclude this literature overview, this chapter presents results from a meta-study on the effectiveness of ALMPs in Europe[4]. In a meta-analysis, based on 137 observations from 95 cross-country studies, Kluve (2006) find that not all ALMPs perform equally well. The author show that most effective among the programs are the private sector incentive schemes (including wage subsidies and start-up grants) and the group of "services and sanctions" (including job search assistance, monitoring and sanctions). Evaluations based on these programs are 40-50 percent more likely to estimate positive treatment effects than evaluations of training programs. In contrast, the author finds that the public sector employment programs (direct job creation) appear to have detrimental effects. As compared to training, evaluations of these programs are 30-40 per cent less likely to estimate positive treatment effects. Also programs targeted at the youth unemployed seem to be less effective than the programs targeted at adults.

Based on these results it can be concluded that the private sector employment programs perform relatively well.

Chapter 4: Methodology

This chapter discusses some methodological issue that arise with the evaluation of active labor market programs. Section 4.1 describes the main evaluation problem. Section 4.2 discusses the adopted estimation strategy.

[4] See Greenberg et al. (2003) for a meta-analysis of US government sponsored training programs and Puerto (2007) for a meta-analysis of youth employment programs.

4.1 The evaluation problem

Assessing the impact of any intervention requires making an inference about how program participants would have performed in the labor market had they not received the treatment. The framework that is used for the empirical analysis of this problem is the potential outcome approach, also known as the Neyman-Roy-Rubin model (Neyman, 1935; Roy, 1951; Rubin 1974)[5].

The main pillars of this model are individuals, treatment and potential outcomes.

According to this approach each individual has two potential outcomes – Y_1, when receiving some treatment D and Y_0, without treatment. D is an indicator variable which is equal to one for the participants (D=1) and zero for the non-participants (D=0). Following the notation in Todd (2007), the actual observed outcome for any individual can be written as $Y_i = DY_{1i} + (1-D)Y_{0i}$. The treatment effect for each individual then can be identified by taking the difference between the two potential outcomes Y_1 and Y_0:

$$\Delta_i = Y_{1i} - Y_{0i} \tag{1}$$

An evaluation problem arises however, because each individual can be observed either as participant or non-participant, whereas his two potential outcomes Y_1 and Y_0 can not be observed simultaneously, i.e. Δ_i is an unobserved random variable, which can not be measured. To estimate Δ_i the researcher needs to find an approximation of the unobserved, counterfactual outcome.

Since individual gains from treatment can never be estimated with confidence the empirical literature concentrates rather on estimating average gains from treatment (Caliendo and Hujer, 2005). The relevant parameter of interest for this analysis is the average treatment effect on the treated (ATT)[6]:

$$ATT = E\,(\Delta | D=1) = E\,(Y_1 - Y_0 | D=1) = E\,(Y_1 | D=1) - E\,(Y_0 | D=1) \tag{2}$$

where D=1 indicates treatment, D=0 indicates no treatment and $E\,(Y_0 | D=1)$ is the expected outcome of the treated had they not been treated.

[5] For an extensive discussion see Heckman et al. (1999), Heckman et al. (1998a), Caliendo and Hujer (2005).
[6] See Todd (2007) for an overview of alternative parameters.

ATT indicates how much on average the individuals who receive treatment benefit from it as compared to a hypothetical situation without treatment. Identifying this effect however, is a complex process and requires some additional assumptions. Experimental and non-experimental studies differ with respect to the assumptions they make (see Smith and Todd, 2005; Smith, 2000; Heckman et al. 1999). In experimental studies (under certain assumptions) randomization will assure that in sufficiently large samples treatment and control groups have on average the same distribution of observable and unobservable characteristics. Therefore the outcome of the randomized-out controls can directly be used as an approximation for the potential outcome of the treated, i.e. $E(Y_0 | D=1) = E(Y_0 | D=0)$, yielding an unbiased estimate of ATT. In studies with non-experimental design however, assignment to treatment is not random. Program participants might be a selected group that differs from non-participants in a non-random way. Consequently, even in absence of the program both groups will have different outcomes, i.e. $E(Y_0 | D=1) \neq E(Y_0 | D=0)$. Estimating ATT by the difference in means of participants and non-participants will lead to a selection bias (Caliendo and Hujer, 2005). In the empirical literature different econometric techniques are suggested to deal with the issue of selection bias[7].

The identification strategy adopted in this analysis is propensity score matching.

4.2 Identification strategy

The method of matching is originally suggested by Rubin (1974). The main idea of matching is to create a comparison group, which is similar to the group of participants in all pre-treatment characteristics that simultaneously affect assignment to treatment and potential outcomes (Hujer and Caliendo, 2000). The crucial assumption behind matching is that conditional on a set of observable characteristics X, potential outcomes (Y_0, Y_1) are independent of treatment status (D). In the literature this assumption is known as the conditional independence (or unconfoundedness) assumption.

Following the notation in Heckman et al. (1998) this assumption can be written as:

Assumption 1 Unconfoundedness: $Y_0 \perp\!\!\!\perp D \mid X$,

[7] For an overview of alternative estimation strategies see Blundell and Costa Dias (2002), Caliendo and Hujer (2005), Heckman et al. (1999), Todd (2007), Bryson et al. (2002).

where $\perp\!\!\!\perp$ denotes independence. Since ATT is the parameter of interest, it suffices only to assume that Y_0 is independent of D, conditional on X. Furthermore, the authors show that to construct the counterfactual outcome and identify ATT an even weaker conditional mean independence assumption on Y_0 suffices:

$$E\,(Y_0 | X, D=1) = E\,(Y_0 | X, D=0) \qquad (3)$$

When this condition holds the counterfactual outcome of participants with observable characteristics X can be constructed from the average outcome of the non-participants with the same observable characteristics X. In other words, treatment becomes random conditional on X. As Caliendo (2008) notes however, this is a strong assumption which requires justification on a case-by-case basis. In order for assumption 1 to hold, all factors that simultaneously affect treatment and potential outcomes must be observed. (The plausibility of this assumption is discussed in the subsequent sections).

Additionally, it is also assumed that for all X the probability of treatment must be smaller than one, i.e.

Assumption 2 Weak Overlap: $Pr(D=1 | X) < 1$

This assumption guarantees that individuals with the same values of X have positive probabilities of being both participants and non-participants (Caliendo and Kopeinig, 2008). In other words, assumption 2 rules out X being a perfect predictor of participation. Therefore, it guarantees that for each participant a counterpart from the non-participants group can be found.

Under assumption 1 and 2 the mean impact of treatment on the treated can be written as

$$ATT = E\,(Y_1 | X, D=1) - E_x[\,E\,(Y_0 | X, D=0) \mid D=1\,], \qquad (4)$$

where $E\,(Y_1 | X, D=1)$ can be estimated from the treatment group and $E_x[\,E\,(Y_0 | X, D=0) \mid D=1\,]$ from the matched on X comparison group (see Smith and Todd, 2005 or Caliendo and Hujer, 2005).

Matching might become however difficult to implement when the set of conditioning covariates is large, the so-called “curse of dimensionality” (Smith and Todd, 2005). In this respect Rosenbaum and Rubin (1983)[8] suggest the use of balancing scores to overcome this dimensionality problem. The authors show that when non-treatment outcomes (Y_0) are independent of treatment status (D) conditional on X, they are also independent of treatment status conditional on a balancing score. The balancing score that is applied in this analysis is the propensity score, i.e. the probability to participate as a function of X, Pr(D=1| X) = P(X). Hence, when assumption 1 and 2 are satisfied also the following equation should hold:

$$E\,(Y_0 | P(X), D=1) = E\,(Y_0 | P(X), D=0), \tag{5}$$

where P(X) stands for propensity score. Conditional on P(X) the expected mean outcome of the treated had they not been treated is the same as the expected mean outcome of the non-treated.

The matching estimator

To exploit assumption 1 or its implication, the conditional mean independence assumption, several matching estimators have been proposed in the literature. Following the notation in Caliendo (2008), a typical matching estimator can be expressed in the form:

$$\Delta^{MAT} = \frac{1}{N_1}\sum_{i\in I_1}[Y_{1i} - \sum_{j\in I_0} w(i,j)Y_{0j}], \qquad \text{with} \quad w(\text{i},1)+\ldots+w(\text{i},\text{N}_0) = 1 \tag{6}$$

where I_1 denotes the set of program participants and I_0 the set of non-participants. N_1 is the number of observations in the set I_1 and N_0 is the number of observations in the set I_0. w indicates the weights given to the *j-th* observation from the control group in constructing the counterfactual for the *i-th* observation from the treatment group (Caliendo, 2008). w depends on the distance between the propensity score of observation i and the propensity score of observation j. The more distant they are, the lower weight is attached to observation j. Hence, the match for each participant $i \in I_1$ is constructed as a weighted average over the outcomes of non-participants (Smith and Todd, 2005). Main difference between the various estimators is that they give different weights to members of the control group in constructing the counterfactual outcome[9]. Some of the estimators, i.e. nearest neighbour, use only a few non-treated individuals to construct the counterfactual outcome of a treated individual, while other

[8] Cited in Smith and Todd (2005).
[9] See Smith and Todd (2005) or Caliendo and Kopeinig (2008) for an overview of matching estimators.

estimators, i.e. kernel matching, use nearly all observations in the control group to construct counterfactual outcomes. As Caliendo and Kopeinig (2008) argue, the choice of matching estimator should be relatively unimportant in large samples. According to the authors, as the sample size grows all estimators become closer to comparing exact matches and hence they should yield the same results. In small samples the situation can be different however, where usually a trade-off arises between bias and variance. With all these considerations in mind, and also following Van den Berg et al. (2008) and Caliendo (2008), this analysis focuses on the kernel matching estimator (KM). KM is a non-parametric matching estimator. A major advantage of KM is its lower variance (achieved by using more observations in constructing the counterfactual outcome) compared to other estimators. Since my treatment and control groups are relatively small, resp. 231 and 1,355 observations, KM might be preferable[10]. Moreover, as Caliendo (2008) states, a second advantage of KM is that bootstrapping is a valid method to draw inference[11]. A drawback of KM is that some of the observations are probably bad matches, resulting in biased estimates. To address these concerns however, a common support condition is imposed, which assures that units from I_1 are matched with units from I_0 only in a region of common support.

In terms of equation (6), the weighting function *w* of the kernel estimator is given by (see Smith and Todd, 2005 or Heckman et al., 1998a):

$$W(i,j) = \frac{G((P_j - P_i)/a_n)}{\sum_{k \in I_0} G((P_k - P_i)/a_n)} \tag{7}$$

where *G(.)* is the kernel function, a_n is a bandwidth parameter and *P* is the estimated propensity score. As noted already, the weights depend on the distance between each unit from the set I_0 and the treated unit for which the counterfactual is being created.

Finally, assumptions have to be made about the choice of kernel function, *G(.)*, and bandwidth, a_n. Caliendo (2008) argues that the choice of kernel function is relatively unimportant. What more important according to the author is, is the choice of a_n. In this respect, Caliendo and Kopeinig (2008) discuss the following trade-off arising with the choice

[10] Later it will become evident that the choice of matching estimator is relatively unimportant in this setting. The different matching algorithms produce nearly identical results.
[11] Bootstrapping refers to calculating standard errors that are adjusted for the additional variability introduced by the estimation of the propensity score and by the process of matching (Sianesi, 2008).

of bandwidth. Small bandwidth results in a small bias and a large variance, while large bandwidth results in a large bias and a small variance. For the choice of bandwidth this analysis follows a rule-of-thumb and uses bandwidth 0.06. The applied kernel function is Epanechnikov, following Van den Berg et al. (2008)[12].

Chapter 5: Data description

This chapter provides information about the dataset. Section 5.1 describes the data and the construction of participants and non-participants samples. Section 5.2 discusses the variables that are used for the estimation of the propensity score.

5.1 Data description and variables

The dataset used in this study combines administrative and survey data, collected in 2007 by the Ministry of Labor and Social Policy (MLSP) in Bulgaria. The data are derived from two sources – the register of the Employment Agency and a follow-up survey. The sample contains information on 7,600 individuals.

In 2007 the MLSP initiated an assessment of the labor market situation of individuals who had previously participated in various labor market programs. For this purpose they selected two samples of 3,800 individuals each. The first sample included individuals who were registered as unemployed and started an active labor market program in the first quarter of 2005, i.e. the participants. The second sample included individuals who were registered as unemployed in the first quarter of 2005, but who did not start any program over this period, i.e. the non-participants[13]. Individuals in both samples were drawn equally form each of the nine main administrative regions in Bulgaria. Consequently, the selected individuals were interviewed in July and August 2007. The interview contained a number of (retrospective) questions concerning issues such as health, education, employment, earnings, etc. In what follows a short description is given of both, the administrative and survey part of the dataset.

The administrative data contain information on age, gender, education, profession, ethnicity, disabilities, place of residence, information of whether an individual is a lone parent or

[12] Sensitivity analysis (with respect to different kernel functions and bandwidths) shows that the estimated results are not sensitive to the choice of kernel function and bandwidth. Results are reported in the subsequent chapters.
[13] It should be noted however, that the non-participants might participate after this quarter.

mother of children younger than 3 years etc. All these variables come from the administrative records of the Employment Agency, as they were recorded in the first quarter of 2005. Additionally the unemployment duration of each individual is known. It is measured from the month of registration with the Employment Agency until the first quarter of 2005.

The follow-up survey contains detailed information on individual's health condition, labor market situation, type of employment, type of contract, number of hours worked, earnings, employer related information, family status, housing, social benefits received etc. For the participants the survey is also informative about the duration of the current employment spell and whether individuals still work for the same employer they used to work for during the program. Unfortunately, there is no information on employment histories or earnings before 2005. Although the survey is quite detailed, a main limitation is that it uses two separate questionnaires, whereas the questions are not always exactly the same[14]. Therefore, this analysis uses only those variables that come from the same questions and from the administrative database of the Employment Agency.

Samples construction

To construct samples of participants and non-participants first, all individuals were excluded who did not respond the survey questionnaires. This procedure resulted in a sample of 2,463 participants and 2,560 non-participants. Further, the sample of participants was divided in sub-groups depending on the type of labor market program they enrolled in. This procedure yielded seventeen program groupings[15]. Then the same procedure was repeated for the non-participants group. The sample of non-participants was divided in seventeen sub-samples depending on whether individuals were eligible for a given program or not[16]. This procedure assures that participants are compared only to non-participants who meet the same eligibility requirements, hence non-participants who potentially could have participated in a given program. Unfortunately, the final number of observations turned to be insufficient for most of the programs. Therefore this analysis examines only subsidized employment programs for long-term unemployed, which is the program with the highest number of observations. Also

[14] Heckman et al. (1998a) find that one of the conditions for matching estimators to perform well is that the same data source (i.e. the same survey) is used for both participants and non-participants so that the various variables are measured in an analogous way.

[15] It was not possible to combine various programs in one group, because most of the programs are targeted at different unemployed and have specific participation criteria.

[16] Some of the individuals turned to be eligible for more than one program, hence they were included in multiple groups.

this program seems to be of major importance as the long-term unemployed make up for a substantial part of total unemployment.

Following Lechner (1999), to avoid dealing with issues such as early retirement and school leaving, this analysis considers only individuals aged between 24 and 55. Additionally, all individuals who had previously participated in active labor market programs are excluded. The remaining 231 participants and 1,335 non-participants constitute the final sample.

Table A-1 presents sample means of the most relevant variables (see Table A-1 in the Appendix). A first glance at Table A-1 shows that participants are on average older, less educated, healthier, less likely to have small children and more likely to be of Bulgarian origin. Also participants have on average shorter unemployment spells and are less likely to be a recipient of social benefits. When it comes to the position sought, participants are on average more likely to accept any job as compared to non-participants.

Outcome variable

The key variable for this analysis, the labor market status in August 2007, is measured by the answer to the question "What is your current employment status?" Based on the answer of this question a binary measure is constructed, indicating the employment state of the individuals. All individuals that hold regular, non-subsidized jobs or are self-employed are considered as employed. Similarly, as non-employed are considered individuals who are unemployed, participants in an active labor market program, student or out of the labor force. Consequently, their outcomes are recorded with "1", respectively "0" in the database. Hence, the impact of the program is assessed about 6 months after the end of the program.[17]

5.2 Is it plausible to assume conditional mean independence?

As Smith and Todd (2005) argue, the success of matching clearly depend on the available data at hand. Matching is a data "hungry" process that requires a lot of detailed information. Moreover, as matching crucially relies on the conditional mean independence assumption, which is in general a non-testable assumption, the choice of covariates to be included in the

[17] Recall that the program started in the first quarter of 2005 and has duration of 24 months.

propensity score should be guided by economic theory, previous research and features of the program (see Smith and Todd, 2005 and Bryson et al., 2002).

Following the argumentation in Bryson et al. (2002), in the estimation of the propensity score should be included only pre-treatment variables that influence simultaneously the decision to participate and the outcome variable. Obviously, the first group of variables that comes to mind are the socio-demographic variables. Previous research shows that characteristics as age, gender, marital status, children, ethnicity and health restrictions are a major category and as such should be included in the propensity score (see Van den Berg et al., 2008; Larsson, 2002; Aakvik, 2001). Therefore, all these variables are included in the specification. The variable children is a binary measure and indicates the presence of children younger than 3 years. In addition to the above socio-demographic characteristics, the specification includes also a variable indicating whether an individual is a lone parent or not. The second group of variables, that are highlight by both economic theory and previous research, are the so-called human capital variables (see Caliendo, 2008). The available attributes in this respect are school degree and profession. The administrative part of the data provides exact information about school degree and name of the followed study. Also, for each individual the exact profession is known[18]. Education and profession can be important factors in capturing unobservable characteristics such as ability. At third place, previous research points the importance of employment histories and labor market dynamics in determining treatment and outcomes. As Heckman et al. (1999) point, including information about unemployment dynamics and labor market histories should capture effects caused by unobservable factors and reduce bias due to unobservables. In this respect, the available characteristics are unemployment duration, previous participation in labor market programs, social benefits received, occupational group and job wanted. Unfortunately, there is no available information about earnings or duration of previous employment spells. However, factors as unemployment duration and social benefits can play an important role in capturing unobservables such as motivation and personal attitudes. Everything else equal, more motivated individuals are expected to have shorter unemployment spells. Also, considering that the social benefits system in Bulgaria is quite tight, individuals who rely on social benefits are expected to have poor personal attitudes and appearances. Moreover, all

[18] The exact profession was recorded as a 4-digit code. After recoding it I created 3 binary measures indicating whether an individual is low, mid or high qualified. I tried also alternative definitions, but they did not make much difference for the results.

individuals in the sample are continuously unemployed for more than 12 months. Practically this means that no one can be a recipient of unemployment benefits.
The characteristics discussed so far are related only to the individuals themselves. However, the assignment to treatment process is not only a process of self-selection. As Aakvik (2001) states, this process is a combination of self-selection, selection by case-workers and first-come, first-serve basis. This is particularly true for the program under examination. Not only individuals determine whether to participate or not, but also case-workers. There is also a random component, such as first-come, first-serve. While we should not worry about the random component, we should control for selection by case-workers. In this respect, Sianesi (2008) suggests including in the propensity score variables that capture observable and unobservable aspects of local employment offices. Following Sianesi (2008), an additional variable is constructed that capture unobserved aspects of local labor offices. The variable local program rate is given by the number of participants in all labor market programs as a proportion of all registered unemployed at a given municipality. As Sianesi suggests, this variable should provide information of local program capacity. Together with this variable two additional variables are included that control for local and regional characteristics. The first variable is the local unemployment rate. The local unemployment is expected to affect both the probability that unemployed people find a job and the chance that they enter a labor market program. Finally, seven regional dummies are included in the specification to control for regional specific effects.

The main question is: is it plausible to assume conditional mean independence based on the data at hand? Probably (even after conditioning on all these variables) there are still remaining differences between treated and non-treated. However, it is not expected that these differences are so big that the estimated results can be invalidated. Later a special attention will be devoted to the issue of hidden bias and how important it can be for the estimated results.

Chapter 6: Implementing the estimator

The first part of this chapter discusses the estimation of the propensity score. Matching details are presented in the second part.

6.1 Propensity score estimation

As discussed already, participants are matched with non-participants based on their probabilities of participation. Since the probabilities are not know beforehand, they are estimated as a function of the covariates discussed in the previous chapter. Table 2 reports results from the probit model, which is used to estimate the probability of participation. Looking at the socio-demographic characteristics it seems that most of the covariates significantly affect the probability of participation. Being older, married, healthier, not having small children and not being single parent are all characteristics that significant increase the probability of participation. Being native Bulgarian is also positively associates with participation, though the variable is significant at 10 percent level. Everything else equal, having a primary, secondary or higher school degree decreases the probability of participation as compared to not having or having a basic degree. Ironically, individuals with higher unemployment spells prior treatment are also less likely to participate. Finally, higher unemployment and program-to-unemployment rates in local districts also significantly increase the probability of program participation.

Based on the reported probit estimates in Table 2, no clear selection pattern can be detected. From one side, "favourable" characteristics such as not-disable, native and shorter unemployment duration are overrepresented among the treated. On the other side however, also "unfavourable" characteristics such as low educational degree are overrepresented in the participants group. Hence, the emerged selection pattern is not consistent with either the hypothesis of cream-skimming, i.e. selecting the most successful, or the hypothesis of bottom fishing, i.e. selecting the most needy[19].

The results of the propensity score estimation are not further discussed in detail, since they are used only to reduce the dimensionality problem.

[19] Aakvik (2001) finds that the individuals selected in the Norwegian vocational rehabilitation programs are those that are most likely to be employed even without the program.

Table 2: Propensity Score Estimation (Probit model)

Variable	Coefficient
Socio-demographic characteristics	
Age (*Ref. 24-29 years*)	
30-44	0.621*** (0.14)
44-55	0.348** (0.149)
Female	-0.037 (0.104)
Married (or cohabiting)	0.235** (0.118)
Children < 3 years (yes)	-1.074*** (0.32)
Single Parent (yes)	-0.961** (0.439)
Disabilities (yes)	-1.103*** (0.247)
Non-native	-0.25* (0.137)
Qualification variables	
School Degree (*Ref. Basic or no Degree*)	
Elementary	-0.448** (0.22)
Secondary	-1.035*** (0.283)
Higher	-1.104*** (0.33)
Job Qualifications (*Ref. Low Qualified*)	
Mid Qualified	0.287 (0.505)
High Qualified	0.271 (0.522)
Occupational group (*Ref. Agriculture*)	
Manufacturing	-0.272 (0.524)
Administration	-0.342 (0.471)
Labor market variables	
Unemployment Duration (in months)	-0.015*** (0.002)
Social Benefits Received (yes)	-0.114 (0.155)
Position Sought (as qualifications)	-0.009 (0.189)
Regional variables	
Regional dummies[1]	
Local Unemployment Rate	0.038** (0.017)
Local Program-to-Unemployment Rate	7.644** (3.838)
Constant	-0.369 (0.342)
Pseudo R-squared	0.26
Log-Likelihood	-480.05
Chi-squared (p-value)	0.0000

Note: Standard errors in parenthesis. Statistical significance at 1, 5 and 10 percent level indicated by ***, **, *.
(1) Additional variables include 7 regional dummies – estimations are excluded due to space considerations.

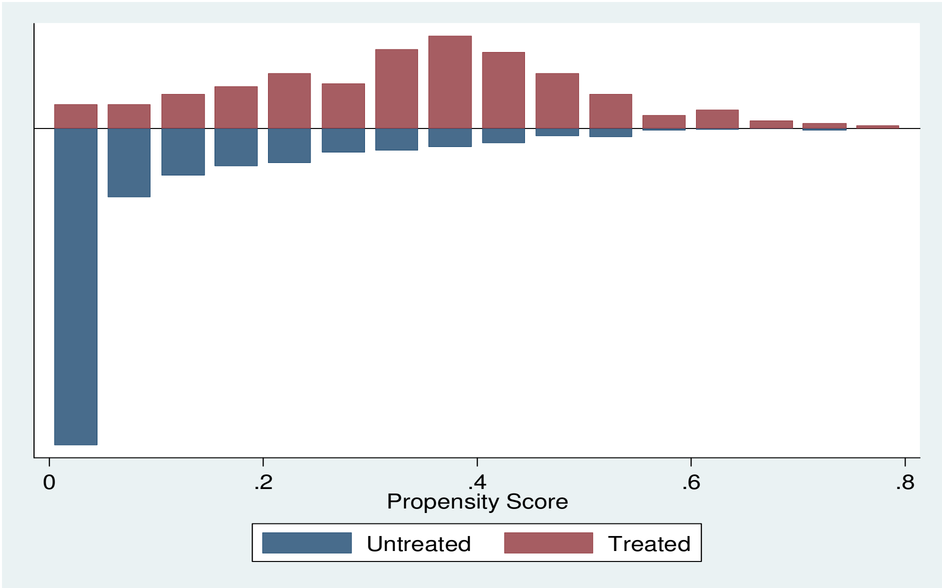
6.2 Propensity score distribution and matching details

Determining the region of common support

After the propensity scores have been estimated, it is essential to examine their distribution. Figure 1 depicts the distribution of the propensity scores for the treated and non-treated. A first glance at the graph shows that in general the overlap between treated and non-treated is satisfactory. Nevertheless, there are some regions, starting around .06, where the distribution of the propensity scores is quite thin, especially in the comparison group. As discussed already in section 4.2, the kernel method uses nearly all observations form the non-treatment group to construct counterfactual outcomes. That is why, to avoid bad matches, it is important to properly determine the region of common support and perform matching within this region. The region of common support is typically determined by discarding all observations whose propensity score is smaller than the smallest and larger than the largest propensity score in the opposite group, i.e. the "min-max criterion". This procedure yields a common support region between [.00086598, .75939682]. Looking at the graph however, it is evident that there are some parts of the distribution just before .75 where the propensity scores in the untreated group are near to zero, even though they fall in the common support. Therefore, also a second stricter imposition of the common support requirement is applied. To ensure that all densities are strictly greater than zero, following Smith and Todd (2005), treated observations whose propensity scores lie in that region are excluded. Hence, the region of common support is determined as those values of the propensity score that have positive density, both in the treatment and the non-treatment group (see Smith and Todd, 2005 for details)[20].

[20] Sensitivity analysis revealed that the estimated ATT is not sensitive to the choice of criterion to impose the common support condition.

Figure 1: Propensity score distribution



Note: Propensity scores of the treated are depicted in the upper half, propensity scores of the untreated in the lower half. The region of common support is [.00086598, .75939682]. The propensity scores are based on the estimations in Table 2.

It should be clear however, that the ATT cannot be estimated for those individuals whose propensity scores lie outside the common support region. In this respect, Bryson et al. (2002) argue that when many individuals are discarded due to imposing the common support this can pose some problems. The estimated treatment effect is then no longer representative for the whole sample, but rather for a sub-population of the sample for which there is support in the non-treatment group. In the present setting though, this seems to be a minor concern. Imposing the common support results only in few individuals being discarded.

Matching details

Before presenting the matching results in chapter 7, this section examines also the quality of matching, i.e. if matching is able to balance the separate covariates. In this respect, DiPrette and Gangl (2004) suggest using the standardized mean difference (SMD) between treatment and control samples as a convenient way to test for covariate balance. The SMD is defined for each covariate as the difference of the sample means in the treated and non-treated groups as a percentage of the square root of the average of the sample variances in both groups (Rosenbaum and Rubin, 1985)[21]. As DiPrette and Gangl note, the standardized mean difference is a simple way to quantify bias between treatment and control samples. The SMD

[21] Cited in DiPrette and Gangl (2004).

test is applied in a number of empirical works (see Bonjour et al., 2001; Sianesi, 2003 and Caliendo, 2008). Table 3 reports results from this test. As can be seen from the table, the mean standardized difference (MSD), which is calculated as an un-weighted average of the standardized differences of all covariates, is substantially reduced after matching. From slightly over 25 percent before matching MSD is reduced to below 3 percent after matching. Although there is no strict rule about how much MSD should be after matching, the reported results are considered as satisfactory[22]. In addition, following Sianesi (2003) and Caliendo (2008), Table 3 reports also two additional indicators – pseudo-R^2 and results from the Likelihood-ratio test. Both authors suggest to re-estimate the propensity score on the matched sample and to compare the resulted R^2 with the R^2 before matching. After matching the covariates should have no explanatory power. As can be seen from the table, pseudo-R^2 after matching is quite low. Also the p-value of the Likelihood-ratio test shows that the joint significance of the regressors is rejected after matching. Based on these indicators it can be concluded that the matching results are satisfactory, hence bias due to covariate imbalance is minimized. Here should be clear however, that the test statistics reported in Table 5 do not provide any indication as of whether the conditional mean independence assumption is satisfied or not (for a further discussion see Bryson et al., 2002).

Table 3: Matching quality

Indicator	Value
MSD – Before Matching	25.24
MSD – After Matching	2.95
Pseudo R^2 – Before Matching	0.26
Pseudo R^2 – After Matching	0.007
$Pr>\chi^2$ – Before Matching	0.000
$Pr>\chi^2$ – After Matching	1.000

Note: Calculations are done using the PSMATCH2 package by Leuven and Sianesi (2003). Before matching results are based on the whole sample; after matching results are based on the matched sample - kernel (Epanechnikov) matching with common support and bandwidth 0.06.

[22] In comparison Bonjour et al. (2001) report MSD after matching between 6 and 10 percent for various programs; Sianesi (2004) reports MSD between 1 and 3.6 percent, Caliendo (2008) reports MSD between 2.39 and 4.77.

Chapter 7: Results

This chapter presents the main empirical results from matching. Section 7.1 discusses the matching estimates. The sensitivity of the results with respect to different kernel functions, bandwidths and matching estimators is analysed in section 7.2.

7.1 Empirical results

As discussed already the main aim of this paper is to estimate the effects of employment subsidies on the probability of regular employment for the participants. Table 4 presents the estimated treatment effects of the program about two and a half years after program start (i.e. a half year after program end). The overall effect of the program in raising employment among the participants is positive and highly significant. Long-term unemployed who participated in the program have on average 33 percent higher employment rate than their non-participating counterparts. Given that we adjusted for various observable characteristics, this is quite a large effect. This estimate is almost identical to the OLS estimate in Table A-2 (see Table A-2 in the Appendix). According to the OLS estimate, everything else equal, participation in the program increases the probability of employment by some 34 percent. The probit estimate, reported in Table A-2, also points in the same direction, though the estimate is a little bit higher, indicating a positive program effect of 36 percent. All three estimates are significant at 1 percent level.

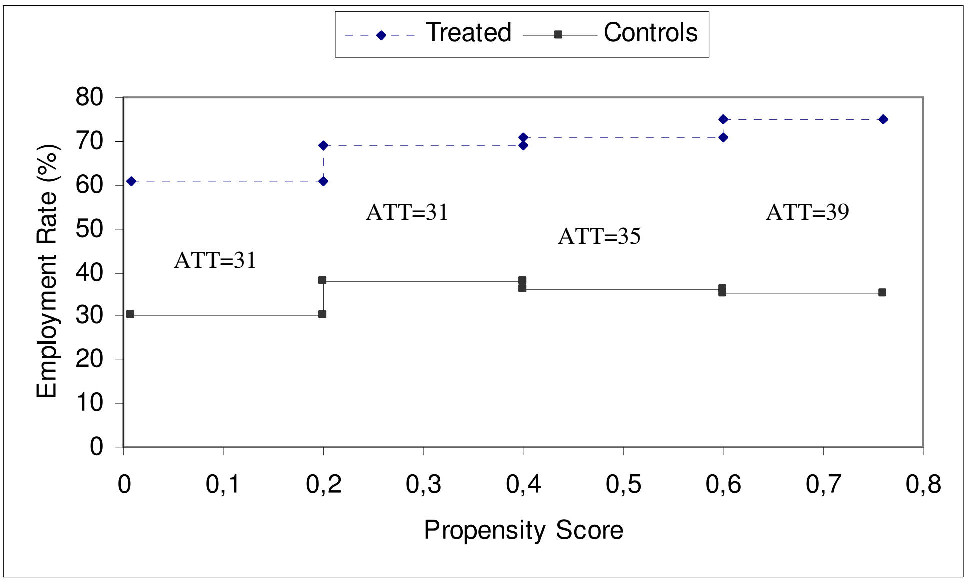
Table 4: Estimation Results – Regular Employment

Outcome variable	Effect	Standard Error	t-value
Employment	0.337	0.044	7.53

Note: Outcome – Regular employment. Kernel (Epanechnikov) matching with common support and bandwidth 0.06. The standard errors are based on 100 bootstrapped replications. Estimations are done using the PSMATCH2 package by Leuven and Sianesi (2003).

Next, the employment effect of the program is also estimated for four different intervals of the propensity score distribution. The first interval includes all individuals with propensity scores below 0.2; the second interval includes all individuals whose propensity scores are between 0.2 and 0.4; the third interval includes individuals with propensity scores between 0.4 and 0.6 and the final interval includes individuals with propensity scores above 0.6. Figure 2 depicts the employment rates in different parts of the propensity score distribution. Individuals with the lowest propensity scores, i.e. the lowest probability of program participation, have on average also the lowest employment rates. About 30 percent of the matched non-participants were employed in 2007, whereas this percentage was about 61 for the group of the participants. This difference is also statistically significant. The estimated average treatment effect on the treated is around 31 percent. The employment rates in the second interval are higher as compared to the first interval. About 38 percent of the matched non-treated and 69 percent of the treated were employed in 2007. The mean effect of the program for the treated is estimated to be around 31 percent for this interval and it is statistically significant at 1 percent level. In the third interval the employment rate is about 71 percent for the group of treated and about 36 percent for the matched non-treated. The average gain from the program for the treated in this interval is abound 35 percent. Finally, individuals with the highest propensity scores (above 0.6), i.e. those that are most likely to participate, gain most from the program. The estimated treatment effect however, is based on a small number of observations and it is not reliable. Overall, the program seems to be successful in raising the probability of employment for all groups.

Figure 2: Treatment effect for different sub-samples



Note: Outcome – Regular employment. Kernel (Epanechnikov) matching with common support and bandwidth 0.06. Calculations are done using the PSMATCH2 package by Leuven and Sianesi (2003).

Though, it should be recognized that our measure of employment is quite conservative. It considers as employed only individuals who are self-employed or employed in regular non-subsidized jobs. This measure of employment can give misleading results to some extend. For example, if there are many people from the group of the non-participants who started a labor market program after the first quarter of 2005 and were still in this program at the time of the survey[23]. From the follow-up survey it is evident that some of the non-participants were participating in subsidized employment programs at the time of the survey. According to our measure of employment all these individuals are considered as non-employed. In that sense, the effectiveness of the program can be overestimated. Therefore, a second outcome variable is constructed, which treats both regular employment and subsidized employment as a success. Table 5 presents the results. The estimated treatment effect is even larger. Participants in subsidized employment programs have on average 34 percent more chance of being employed in a regular or subsidized job than the matched non-participants.

Table 5: Estimation Results – Regular and Subsidized Employment

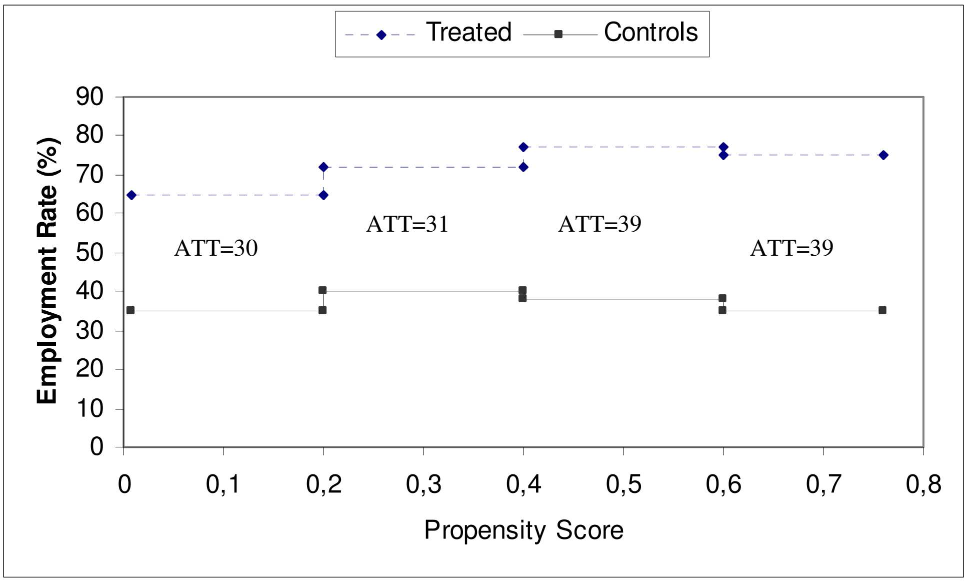
Outcome variable	Effect	Standard Error	t-value
Employment	0.348	0.044	7.86

Note: Outcome – Regular and subsidized employment. Kernel (Epanechnikov) matching with common support and bandwidth 0.06. The standard errors are based on 100 bootstrapped replications. Estimations are done using the PSMATCH2 package by Leuven and Sianesi (2003).

Figure 3 depicts the employment effects of the program for different strata of the propensity score distribution. The program seems to be less effective for individuals with the lowest propensity scores. Participation in the program increases the probability of employment with about 30-31 percent for individuals with propensity scores below 0.4. Individuals with propensity scores above 0.4 enjoy the highest gains from participation in the program. For this group the program is estimated to increase the probability of employment by about 39 percent. From that point, the selection rule of the program appears to be correct.

[23] Remember that as non-participants were defined all individuals who were registered as unemployed, but did not start any program in the first quarter of 2005. However, it is possible that some of them participated in a program after this period.

Figure 3: Treatment effect for different sub-samples



Note: Outcome – Regular and subsidized employment. Kernel (Epanechnikov) matching with common support and bandwidth 0.06. Calculations are done using the PSMATCH2 package by Leuven and Sianesi (2003).

7.2. Sensitivity analysis

Additional analysis reveals that the main estimate (outcome – regular employment) is not sensitive to the choice of kernel function or bandwidth. Table 6 presents the results. The estimates are virtually identical with the baseline estimate of 0.337. Imposing a stricter bandwidth (epanechnikov kernel function with bandwidth 0.01) though, results in a lower estimate. Nevertheless, the estimated effect is close to the baseline and also statistically significant at 1 percent.

Table 6: Different Kernel functions and bandwidths

Kernel function / bandwidth	Effect	Standard Error	t-value
Epanechnikov / 0.01	0.319	0.038	8.23
Epanechnikov / 0.1	0.337	0.039	8.62
Normal / 0.06	0.339	0.038	8.80
Uniform / 0.06	0.335	0.039	8.39
Tricube / 0.06	0.337	0.042	7.87

Note: Outcome – Regular employment. Estimates are done using the PSMATCH2 package by Leuven and Sianesi (2003). The standard errors are based on 100 bootstrapped replications.

Additionally, the baseline estimate also does not seem to be sensitive to the choice of matching estimator. Table 7 presents results from different matching estimators. All matching estimators point to an average treatment effect on the treated of around 33 percent. Based on the results in both tables we can conclude that the estimated effect is not sensitive to the choice of kernel function, bandwidth or matching estimator. Most importantly, in the next section we will examine the robustness of the estimated effect with respect to hidden bias.

Table 7: Estimation results – Regular Employment

Matching method	Effect	Standard Error	t-value
Nearest Neighbor[1]	0.334	0.043	7.62
Caliper Matching[2]	0.334	0.057	5.80
Radius Matching[3]	0.336	0.038	8.74
Stratification[4]	0.337	0.039	8.59

Note: The estimations for the Nearest Neighbour, Caliper and Radius Matching methods are done using the PSMATCH2 package by Leuven and Sianesi (2003). The estimation for the Stratification method is done by using the ATTS.ado by Becker and Ichino (2002). Normal standard errors.

1. Refers to 1-to-1 matching without replacement.
2. Refers to 1-to-1 nearest neighbor matching within a caliper of 0.05.
3. Refers to matching within a radius of 0.05. Differently from the caliper matching, the radius matching uses all control units within the caliper.
4. Refers to matching within intervals of the propensity score. The propensity score is divided into intervals (strata) and for each interval the treatment effect is calculated by taking the mean difference in outcomes between treated and non-treated units. The overall impact is calculated as a weighted average of the interval impacts.

Chapter 8: Selection on unobservable variables

The estimated results crucially rely on the conditional mean independence assumption, which assumes that all factors affecting treatment and potential outcomes are observed. This assumption however, may or may not be satisfied, which is difficult to determine without experimental data at hand. Even though we conditioned on many background variables, still there may be unobservable factors as motivation, ability, preferences etc. that are not fully captured in the data. If these unobserved factors simultaneously affect the decision to participate and the outcome of the individuals, then the estimated results will be biased. In other words, the matching estimator is not robust to hidden bias. The aim of this section is to examine how hidden biases of various magnitudes might alter our conclusions about the effectiveness of the program (see Rosenbaum, 2005).

Following Aakvik (2001)[24], let assume that the probability of program participation is given by:

$$\pi_i = \Pr(D_i = 1| x_i) = F(\beta x_i + \gamma u_i), \tag{8}$$

where x_i is the vector of background variables that we included in the propensity score, u_i is an unobserved variable and γ is the effect of this variable on the probability to receive treatment. Obviously, when γ equals zero the probability to participate is determined only by the observed characteristics x. If γ is different than zero however, two individuals who are exactly matched on their observable characteristics will have different probabilities of receiving the treatment. Assuming that F is the logistic distribution, two individuals *i* and *j* have the following odds of receiving treatment: $(\pi_i/1-\pi_i)=\exp(\beta x_i + \gamma u_i)$ and $(\pi_j/1-\pi_j)=\exp(\beta x_j + \gamma u_j)$. The odds ratio then can be written as:

$$\frac{\pi_i(1-\pi_j)}{\pi_j(1-\pi_i)} = \frac{\exp(\beta x_j + \gamma u_j)}{\exp(\beta x_i + \gamma u_i)} = \exp[\gamma(u_i - u_j)] \tag{9}$$

If matching is successful in balancing x, then after matching the set of observables x should be cancelled out. Then the odds of receiving treatment depend on the parameter γ and the difference in the unobservable variables $(u_i - u_j)$. When there is no differences in unobservable variables, i.e. $u_i = u_j$, or when the unobserved variables do not influence the probability of receiving treatment, i.e. $\gamma = 0$, the odds ratio will equal one. Hence, both individuals will have the same odds of receiving treatment, implying that there is no selection bias due to unobservables. The whole purpose of this sensitivity analysis is to examine how our conclusions about program effectiveness will change if we change the values of $(u_i - u_j)$ and the parameter γ.

Following Aakvik (2001), lets assume that the unobserved variable u_i can take on two values, zero and one, and also that the unobserved variable is motivation. For simplicity reasons it is

[24] If not otherwise stated, the discussion in this section is based on Aakvik (2001).

further assumed that an individual can be either motivated in which case $u_i = 1$ or not motivated in which case $u_i = 0$. Then Aakvik shows that equation 9 implies the following bounds on the odds ratio of receiving treatment:

$$\frac{1}{e^{\gamma}} \leq \frac{\pi_i(1-\pi_j)}{\pi_j(1-\pi_i)} \leq e^{\gamma} \qquad (10)$$

Clearly, when $e^{\gamma} = 1$, which is the case in experimental studies, then treated and matched non-treated will have the same probability of participation. If for example, $e^{\gamma} = 2$ then two individuals with the same x will have different probabilities of participation, i.e. one of them will be twice as likely as the other to receive the treatment (Rosenbaum, 2005). As Rosenbaum (2005) notes, e^{γ} can be seen as a degree of departure from random assignment. e^{γ} is however, an unknown parameter in studies with non-experimental design. What the sensitivity analysis does in that case is to try several values of the parameter to see how the conclusion might change (Rosenbaum, 2005). A study is said to be sensitive to hidden bias when even small departures from random assignment (i.e. e^{γ} close to one) change the conclusion about the effectiveness of the program.

Aakvik (2001) shows that the Mantel-Haenszel (1959) test can be used to test for no treatment effects in case of binary outcomes.

Table 9 shows the results from the Mantel-Haenszel (1959) test for different values of e^{γ}. The two bounds in the table are under the assumption that we have underestimated, respectively, overestimated the treatment effect. The p-value (p^{+}) shows the significance level under the assumption that we have overestimated the treatment effect. (Given that we have a highly significant positive treatment effect, it is less interesting to examine the case of underestimation, then the estimated effect would become even more significant for higher values of e^{γ}). Starting from 1, i.e. no hidden bias, the value of e^{γ} is gradually increased until it changes inference about the estimated treatment effect (Becker and Caliendo, 2007). In this way it is possible to assess how strong unmeasured influences have to be so that the estimated positive treatment effect would have arisen purely due to unobserved positive selection into the program, i.e. those that are more likely to participate are also more likely to be employed[25]. As can be seen from the table the critical value of e^{γ} at which the estimated treatment effect becomes sensitive to hidden bias is between 2.75 - 3. This means that, given the same x, if there is an unobserved factor that makes one of the two groups three times more likely to participate than the other and also this unobserved factor is almost perfectly correlated with employment, then the estimated positive treatment effect can be questioned. It should be clear though, that the test does not say whether there is such unobserved factor or not. Merely, it says that the confidence interval for the estimated effect would include zero if treated and matched non-treated differ by a factor of 3 in their odds of receiving treatment (Becker and Caliendo, 2007).

Table 9: Mantel-Haenszel (1959) test for hidden bias

e^{γ}	Bounds	Critical p-value (p^{+})
1	7.09**	6.7e-13
1.25	5.91 – 8.30**	1.7e-09
1.5	4.96 – 9.30**	3.5e-07
1.75	4.16 – 10.17**	0.000016
2	3.47 – 10.93**	0.000254
2.25	2.87 – 11.61**	0.00202
2.5	2.33 – 12.23**	0.009676
2.75	1.85 – 12.80**	0.031798
3	1.41 – 13.32	0.078627
3.25	1.009 – 13.81	0.156336
3.5	0.63 – 14.27	0.262624
3.75	0.28 – 14.70	0.387071
4	-0.03 – 15.10	0.515469

Note: Assumption lower bounds: underestimation of treatment effect. Assumption upper bounds: overestimation of treatment effect. Value p^{+}: significance level under the assumption of overestimation of treatment effects. ** indicates that the estimated treatment effect is not sensitive to selection bias at 5 percent level. The estimates are done using the MHBOUNDS.ado by Becker and Caliendo (2007).

In comparison to other studies, our result seems to be quite robust to possible deviations from the no-bias assumption. For example, Aakvik (2001) finds that most of his estimates are

[25] As Aakvik (2001) notes, it should be made clear distinction between selection into the program based on observables (which we discussed in the previous sections) and selection on unobservables, given the same vector x of observables.

sensitive to hidden bias already at $e^{\gamma} = 1.25$. DiPrete and Gangl (2004) find that their estimated treatment effects become sensitive at e^{γ} respectively 1.15, 1.60 and 2.30 for different outcome variables. Caliendo et al. (2005) report critical values of e^{γ} ranging from 1.20, respectively 1.30 for the sample of men and women in East Germany to 1.55, respectively 1.80 for the sample of men and women in West Germany.

In that sense, the estimated in this study treatment effect of 0.337 seems to be relatively insensitive. Here should be stressed however, that this test is by no means a justification of the conditional mean independence assumption. Rather, it is an examination of how departures from this assumption would impact on the estimated effect.

Chapter 9: Concluding remarks

This chapter summarizes the main findings of the paper. Upon these findings the research question will be answered and conclusions will be drawn about the effectiveness of the studied program. Recommendations for further research are given at the end of the chapter.

9.1 Research question

In the introduction of the paper the following research question was formulated:

Are subsidized employment programs an effective instrument to bring long-term unemployed workers back to work?

To answer this question in a systematic way the paper was organized in chapters. First, we started with a short description of the institutional framework in Bulgaria. A concise discussion was provided of both passive and active labor market policies. Also main features of the program under examination, subsidized employment for long-term unemployed, were analysed. Next, the paper presented a short review of previous empirical studies that examine the effectiveness of employment subsidies. The main emerging conclusion from this review was that in general employment subsidies exhibit positive treatment effects. Further, the paper provided an extensive discussion of some methodological issues that arise with the evaluation of labor market programs. The basic form of the evaluation problem was laid out and also an estimation strategy was suggested to solve the evaluation problem. The paper provided also a description of the dataset used in the empirical analysis. Furthermore, as the propensity scores are not known they had to be estimated on hand of the available pre-treatment variables that simultaneously affect participation in the program and the outcome variable. Therefore, the paper provided information on how the propensity scores were estimated and also some matching details. The main empirical results from matching were presented in chapter 7. The estimated treatment effect on the treated suggested that treated individuals were on average about 33 percent more likely to be employed in August 2007 than non-treated. Additional sensitivity analysis revealed that the estimated effect is not sensitive to different kernel functions, bandwidths and matching estimators. Finally, the estimated effect was scrutinized to see how sensitive it is to hidden bias. From the sensitivity analysis, carried out in the framework of Rosenbaum bounds, it appeared that the estimated effect is relatively insensitive to possible deviations from the no-bias assumption.

Based on these results it can be concluded that employment subsidies, given to private employers to hire long-term unemployed workers, are an effective instrument to combat long-term unemployment.

9.2 Recommendations

It is fair to acknowledge that the analysis in this paper is based on a dataset that may not include all factors affecting treatment and potential outcomes. Even though we found that the estimated treatment effect is not sensitive to hidden bias, still it is desirable to re-assess the impact of the program when a better dataset is available in the future. Second, further research is needed to examine whether the estimated positive treatment effect of the program is only short lived or it is also sustained in the medium and long-run. Third, considering that Bulgaria is a developing country with limited resource, it is important in future research to consider also the costs and benefits of the program. Finally, before making any policy recommendations for expanding the program, one should also examine the unintended side effects of the program. For example, it could be the case that (as a result of the subsidy) firms substitute regular for subsidized workers, or that firms would have hired the same long-term unemployed workers even without the subsidy. In such case one speaks of *substitution* and *deadweight effects* of the program. Calmfors et al. (2002) find that Swedish employment subsidies in the 1990s caused substantial crowding-out (displacement) effects. The estimated displacement was largest for employment programs that closely resembled regular jobs. This suggests that the presently evaluated program, which almost perfectly resembles regular employment, may have considerable crowding-out effects. Therefore, it is essential to examine possible side effects of the program before implementing it on a larger scale.

References

Aakvik, A. (2001), Bounding a Matching Estimator: The Case of a Norwegian Training Program, *Oxford Bulletin of Economics and Statistics* 63 (1): 115 – 143

Becker, S. and Caliendo, M., (2007), Sensitivity analysis for average treatment effects, *The Stata Journal* 7 (1): 71 – 83

Becker, S. and Ichino, A., (2002), Estimation of average treatment effects based on propensity score, *The Stata Journal* 2 (4): 358 - 377

Berg, G., van den, Bergemann, A. and Caliendo, M., (2008), The Effect of Active Labor Market Programs on Not-Yet Treated Unemployed Individuals, IZA Discussion Papers Series, Discussion Paper no. 3825, Institute for the Study of Labor (IZA), Bonn, Germany

Bergemann, A. and van den Berg, G., (2006), Active Labor Market Policy Effects for Women in Europe: A Survey, IZA Discussion Paper Series, Discussion Paper no. 2365, Institute for the Study of Labor (IZA), Bonn, Germany

Blundell, R. and Costa Dias, M., (2002), Alternative Approaches to Evaluation in Empirical Microeconomics, CEMMAP Working Papers Series, Working Paper no. CWP 10/02, University College London and Institute for Fiscal Studies, London, UK

Bolvig, I., Jansen, P. and Rosholm, M., (2003), The Employment Effects of Active Social Policy, IZA Discussion Papers Series Discussion Paper no. 736, Institute for the Study of Labor (IZA), Bonn, Germany

Bonjour, D., Dorsett, R., Knight, G., Lissenburgh, S., Mukherjee, A., Payne, J., Range, M., Urwin, P. and White, M., (2001) New Deal for Young People: National Survey of Participants: Stage 2, Employment Service Research and Development Report ESR67, Employment Service, Sheffield, UK

Bryson, A., Dorsett, R. and Purdon, S., (2002), The use of propensity score matching in the evaluation of active labour market policies, DWP Working Papers Series, Working Paper no.4, Department for Work and Pensions (DWP), London, UK

Caliendo, M. and Kopeinig, S., (2008), Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys* 22 (1): 31 – 72

Caliendo, M., (2008), Star-Up Subsidies in East-Germany: Finally a Policy that Works? IZA Discussion Papers Series, Discussion Paper no. 3360, Institute for the Study of Labor (IZA), Bonn, Germany

Caliendo, M. and Hujer, R., (2005), The Microeconometric Estimation of Treatment Effects – An Overview, IZA Discussion Papers Series, Discussion Paper no. 1653, Institute for the Study of Labor (IZA), Bonn, Germany

Caliendo, M., Hujer, R. and Thomsen, S., (2005), The Employment Effects of Job Creation Schemes in Germany: A Microeconometric Evaluation, IZA Discussion Papers Series, Discussion Paper no. 1512, Institute for the Study of Labor (IZA), Bonn, Germany

Calmfors, L., Forslund, A. and Hemström, M., (2002), Does active labour market policy work? Lessons from the Swedish experiences, IIES Seminar Paper no. 700/2002, Institute for International Economic Studies (IIES), Stockholm University, Sweden

DiPrete, T. and Gangl, M., (2004), Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments, WZB (Social Science Research Center Berlin) Discussion Papers Series, Discussion Paper order-nr. SP I 2004-101, ISSN nr. 1011-9523, Berlin, Germany

Forslund, A., Johansson, P. and Lindqvist, L., (2004), Employment subsidies– A fast lane from unemployment to work?, IFAU Working Papers Series, Working Paper no. 2004:18, Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Sweden

Garibaldi, P., Makovec, M. and Stoyanova, G., (2001), *From transition to EU accession: the Bulgarian labor market during the 1990s*, World Bank Publications, ISBN 0821349163

Gerfin, M., Lechner, M. and Steiger, H., (2005), Does Subsidised Temporary Employment Get the Unemployed Back to Work? An Econometric Analysis of Two Different Schemes, *Labour Economics* 12 (6): 807-835

Greenberg, D., Michalopoulos, C. and Robins, P., (2003), A Meta-Analysis of Government Sponsored Training Programs, *Industrial & Labor Relations Review* 57 (1): 31-53

Heckman, J., LaLonde, R. and Smith, J., (1999), The Economics and Econometrics of Active Labor Market Programs, Chapter 31 in: Ashenfelter, O. and Card, D., (Eds) *Handbook of Labor Economics* vol. 3 (1), North-Holland, pp. 1865 - 2097

Heckman, J., Ichimura, H. and Todd, P., (1998), Matching As An Econometric Evaluation Estimator, *Review of Economic Studies* 65 (2): 261 – 294

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

Kluve, J., (2006), The effectiveness of European active labour market policy, IZA Discussion Papers Series, Discussion Paper no. 2018, Institute for the Study of Labor (IZA), Bonn, Germany

Kluve, J., Lehmann, H and Schmidt, C., (1999), Active Labour Market Policies in Poland: Human Capital Enhancement, Stigmatization or Benefit Churning, IZA Discussion Papers Series, Discussion Paper no. 30, Institute for the Study of Labor (IZA), Bonn, Germany

Koning, J. de, Kotzeva, M. and Tzvetkov, S., (2007), Mid-term Evaluation of the Bulgarian Programme "From Social Assistance to Employment", In Koning, J. de, (Eds) *Employment and Training Policies in Central en Eastern Europe: A Transitional Labour Market Perspective*, Dutch University Press, Amsterdam.

Larsson, L., (2002), Evaluating social programs: active labour market policies and social insurance, IFAU Dissertation Series 2002:1, Presented at the Department of Economics, Uppsala University, Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Sweden

Lechner, M., (1999), An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany, IZA Discussion Papers Series, Discussion Paper no. 93, Institute for the Study of Labor (IZA), Bonn, Germany

Leuven, E. and Sianesi, B., (2003), PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Statistical Software Components no. S432001, Boston College Department of Economics, revised version 3.1.5 02 May 2009 E. Leuven, B. Sianesi, http://ideas.repec.org/c/boc/bocode/s432001.html

Mantel, N. and Haenszel, W., (1959), Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease, *Journal of the National Cancer Institute* 22: 719 - 748

OECD, (1998), Labour market and social policy in Bulgaria, seminar held in Sofia on 5 March 1998, Organisation for Economic Co-operation and Development (OECD), Centre for Co-operation with Non-members, Directorate for Education, Employement, Labour and Social Affairs

Puerto, O., (2007), Labor Market Impact on Youth: A meta-analysis of the Youth Employment Inventory, Background Paper for the World Bank's 2007 Youth Employment Inventory

Rosenbaum, P., (2005), Observational Study, In: Everitt, B. and Howell, D., (Eds), *Encyclopedia of Statistics in Behavioral Science* vol. 3, John Wiley & Sons, Ltd, pp. 1451-1462

Rosholm, M. and Svarer, M., (2004), Estimating the Threat Effect of Active Labour Market Programmes, CAM no. 2004-14, Centre for Applied Microeconometrics (CAM), Institute of Economics, University of Copenhagen

Rubin, D., (1974), Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies, *Journal of Educational Psychology*, 66, pp. 688-701

Sianesi, B., (2008), Differential effects of active labor market programs for the unemployed, *Labour Economics* 15 (3): 370 – 399

Sianesi, B., (2003), An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s, IFS Working Papers Series, Working Paper no. 02/01, The Institute for Fiscal Studies (IFS), London, UK

Smith, J., (2000), A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies, *Swiss Journal of Economics and Statistics* 136 (3): 247 - 268

Todd, P., (2008), Matching Estimators, In: Durlauf, S. and Blume, L., (Eds), *The New Palgrave Dictionary of Economics*, Second Edition, Palgrave Macmillan

Todd, P., (2007), Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated, Chapter 60 in: Schultz, P. and Strauss, J., (Eds), *Handbook of Development Economics,* vol. 4, North-Holland, pp. 3847 - 3894

Walsh, K., Kotzeva, M., Dolle, E. and Dorenbos, R., (2001), Evaluation of Net Impact of Labour Market Programmes in Bulgaria, Project, World Bank and the Netherlands Economic Institute (NEI)

World Bank, (2003), *Bulgaria: public expenditure issues and directions for reform*, World Bank Publications, ISBN 0821354000

Zhang, T., (2003), Identifying treatment effects of active labour market programmes for Norwegian adults, Memorandum no. 26/2003, Department of Economics, University of Oslo

Appendix

Table A-1: Selected Descriptives

Variable	Participants	Controls
Number of observations	**231**	**1,335**
Socio-demographic characteristics		
Age (in years)	39.56	37.87
Female	0.64	0.67
Married (or cohabiting)	0.81	0.68
Children < 3 years (yes)	0.008	0.04
Single Parent (yes)	0.004	0.03
Disabilities (yes)	0.01	0.17
Non-native	0.24	0.29
Qualification variables		
School Degree		
Basic or no degree	0.07	0.04
Elementary	0.28	0.20
Secondary	0.54	0.59
Higher	0.10	0.14
Job Qualifications		
High Qualified	0.23	0.28
Mid Qualified	0.38	0.33
Low Qualified	0.37	0.37
Occupational group		
Manufacturing	0.57	0.57
Administration	0.05	0.05
Agriculture	0.36	0.35
Labor market variables		
Unemployment Duration (in months)	24.45	34.68
Social Benefits Received (yes)	0.07	0.16
Job Wanted (as qualifications)[1]	0.64	0.73

Note: Unrounded numbers.

(1) *Job Wanted* is a binary variable. It takes the value of one if a person looks for a specific job, related to qualifications and the value of zero if a person is prepared to take any job.

Table A-2: OLS and Probit estimates - Employment

Variables	OLS	Probit Marginal effects[1]	Matching
Treatment effect	0.345*** (0.034)	0.366*** (0.05)	0.337*** (0.04)
Socio-demographic characteristics			
Age (*Ref. 24-29 years*)			
30-44 years	-0.034 (0.029)	-0.031 (0.026)	
44-55 years	-0.081*** (0.031)	-0.068*** (0.031)	
Female	-0.007 (0.024)	-0.011 (0.022)	
Married (or cohabiting)	0.014 (0.024)	0.011 (0.023)	
Children < 3 years (yes)	0.074 (0.061)	0.057 (0.059)	
Single Parent (yes)	-0.041 (0.058)	-0.027 (0.058)	
Disabilities (yes)	-0.258*** (0.024)	-0.159*** (0.055)	
Non-native	-0.042 (0.029)	-0.041 (0.029)	
Qualification variables			
School Degree (*Ref. Basic or no Degree*)			
Primary	0.094* (0.049)	0.102 (0.061)	
Secondary	0.131** (0.064)	0.156** (0.078)	
Higher	0.224*** (0.078)	0.235** (0.096)	
Job Qualifications (*Ref. Low Qualified*)			
Mid Qualified	-0.12 (0.105)	-0.087 (0.062)	
High Qualified	-0.103 (0.11)	-0.08 (0.066)	
Occupational group (*Ref. Agriculture*)			
Manufacturing	0.174 (0.111)	0.196* (0.135)	
Administration	0.124 (0.095)	0.146 (0.108)	
Labor market variables			
Unemployment Duration	-0.001*** (0.0004)	-0.01*** (0.0005)	
Social Benefits Received	-0.22*** (0.023)	-0.158*** (0.05)	
Job wanted	0.03 (0.042)	0.034 (0.043)	
Regional variables			
Regional dummies[1]			
Local Unemployment Rate	-0.011*** (0.003)	-0.01*** (0.004)	
Constant	0.373*** (0.084)		

# Observations	1,566	1,566
R-squared	0.24	0.22

Note: 1. The marginal effects are estimated at the mean value of the continuous variables and for a discrete change from 0 to 1 for the binary variables. Standard errors in parenthesis. Statistical significance at 1, 5 and 10 percent level is indicated by ***, **, *.

1. Additional variables include 7 regional dummies. They are not reported in the table due to space considerations, results are available on request.