INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

TU WIEN

CONTINUING EDUCATION CENTER

MSc Economics

Evaluating labour market programs using propensity score methods

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by

Selver Derya Uysal

Mária Repková

1127067

Vienna, 8 June 2013

MSc Economics

Affidavit

I, Mária Repková

hereby declare

that I am the sole author of the present Master's Thesis,

Evaluating labour market programs using propensity score methods

46 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, 8 June 2013

Signature

Contents

1 Introduction **1**

2 Theoretical framework **4**

2.1 Basic setup and average treatment effect 4

2.2 Weighting using the propensity score 7

2.3 Doubly-robust estimator 9

2.4 Asymptotic properties of estimators 11

3 Empirical analysis **15**

3.1 Data description 15

3.2 Empirical results 16

4 Conclusion **24**

References **26**

Appendix **28**

List of Tables

1 Number of observations by program type and gender 15

2 Estimated average treatment effects 20

3 Mean difference of outcomes between the treated and the control group 21

4 Covariates 30

5 Estimation of propensity score using probit model 32

6 Matching estimation results 35

List of Figures

1 Histograms of propensity scores estimated for Female - Active job-search programs . . . 18
2 Histograms of propensity scores estimated for Female - Training programs . . . 18
3 Histograms of propensity scores estimated for Male - Active job-search programs . . . 19
4 Histograms of propensity scores estimated for Male - Training programs . . . 19
5 Average treatment effects over time . . . 23

Appendices

Appendix A: Consistency of doubly-robust estimator 28

Appendix B: List of covariates 30

Appendix C: Propensity score estimation 32

Appendix D: Matching estimation results 35

Appendix E: Matlab code 36

Abstract

Estimation of causal effects of participation in a program or treatment on some outcome of interest can play an important role in many areas including economics. In this paper I consider methods for estimation of average treatment effects that assume that treatment assignment is unconfounded with outcomes conditional on the set of relevant covariates. I present the estimators based on weighting by propensity score, the conditional probability of receiving the treatment and the estimator that combines weighting with the regression approach, called a doubly-robust estimator, because it is consistent even if one of the underlying models used for regression or weighting is misspecified. Along with the outlined estimation methods I describe the theoretical properties of the estimators that lead to the asymptotic standard error estimation. I then apply these methods to estimate the average treatment effect of participation in two types of active labour market programs in Austria, active job-search programs and training programs, on the future employment of individuals after four years following the program start. I use an individual administrative data set from the analysis of Hofer, Sellner and Weber (2007). According to the results of doubly-robust estimator active job-search programs perform better than the training programs particularly for women, where estimated effects are significantly positive for both programs. For men a negative treatment effect is estimated in case of training programs. The results suggest the presence of the stronger lock-in effect of training programs comparing to job-search programs, especially for men, that is due to the initial reduction in the job search effort during the training program participation. Finally I look at the average treatment effect estimated in each quarter up to twenty quarters after the program start what provides a further evidence for the later occurrence of the positive treatment effect for training programs compared to the active job-search programs.

1 Introduction

Estimation of causal effects of programs or policies plays an important role in economics and other social sciences. The main interest lies in evaluating the effect of participation of units (individuals, households, states etc.) in a program or in a treatment on some outcome of interest. A crucial feature is that each unit, in principle, can be either exposed to treatment or not, and we observe the outcome only for one situation, nevertheless our interest is in comparison of the two outcomes for the same unit when exposed, and when not exposed, to the treatment. In order to evaluate the effect of treatment we compare the outcomes from two distinct groups, the treated group and the control group. This approach requires an assumption that adjusting treatment and control groups for differences in some observed covariates, pretreatment variables, removes the biases in comparison of the two groups, referred to as unconfoundedness assumption. A popular approach to estimate the causal difference is to adjust for this confounding by using a propensity score, the conditional probability of receiving the treatment. Various methods for treatment effect estimation has been proposed[1] and I will particularly focus on methods using weighting based on propensity score and the combination method using both weighting and regression approach. I will discuss three inverse probability weighted estimators as reviewed in Lunceford and Davidian (2004) along with one estimator that combines weighting with regression, so-called doubly-robust estimator, that essentially provides some robustness against misspecification in a sense that the parameter estimator remains consistent even in case of misspecification of one of the underlying models for the regression or for the weighted method.

In this paper, I estimate the impact of active labour market programs on the future employment of individuals. I use an individual administrative data set from the analysis of Hofer, Sellner and Weber (2007). Data set combines information from Austrian social security records, labour market history and labour market program participation, with personal characteristics of individuals. Following the research of Hofer, Sellner and Weber (2007) I focus on two types of programs,

[1] Imbens and Wooldridge (2008) review the recent development in program evaluation.

active job-search programs and training programs and estimate the average treatment effect of the program participation using propensity score weighting approach. Active labour market programs are generally considered a tool to reduce an unemployment, however differences across the programs raise the question of effectiveness of particular types of programs. Job-search assistance programs are shown to perform better over traditional training programs in several studies such as Kluve (2006) who considers hundred European evaluation studies and shows better outcome of job-search programs on post-program employment rates. Also Weber and Hofer (2004), who use time-of-events method, show that participation in the active job-search programs significantly reduces an unemployment duration compared to the training programs which show a negative effect on employment in the short run. This might be a result of what is considered to be an important drawback of the traditional programs that they lead unemployed individuals to reduce their search effort during participation, referred to as lock-in effect. On the contrary, active job-search programs aim to increase search effort and search efficiency by activating and encouraging the participants to move out of unemployment. Another evidence present in the studies is the male-female difference in the effects of the programs. Lechner and Wiehler (2011) argue that better outcome of active labour market programs for women in terms of increased employment is due to their effect on reduction of share of women leaving the labour force and come with the reduction or postponement of pregnancies.

Hofer, Sellner and Weber (2007) analyse the two groups of programs, active job-search programs and training programs, separately for males and females and by estimating the average treatment effect on treated provide evidence that in the medium run (4 years) active job-search programs are more effective than the traditional training programs. However, for both types of programs, a significant effect on post-employment was found only for the female participants. Using an alternative method of propensity score weighting and a doubly robust method I estimate the average treatment effect of participating in programs on a future employment of individuals after four years following the program start. Analyzing the two program types separately for females and males I find, in accordance with the previous findings, that active job-search programs perform better than training programs. For female participants are the effects of both programs significantly positive, however bigger for the active job-search programs when following the results from doubly-robust estimation. On the contrary participating of males in training programs results in the significantly negative treatment effect. Estimated average treatment effect results suggest the higher effectiveness of active job-search programs compared to training programs for female participants and support the assumption of strong lock-in effect of training programs in the medium run for male participants.

In Section 2, I discuss a basic framework of potential outcomes, that allows to define a causal effect, or treatment effect, and assumptions necessary for adjustment for confounding. I describe the average treatment effect estimation under unconfoundedness using weighting approach based on propensity score and the combination of weighting and regression approach and outline the theoretical properties of the estimators that allow for asymptotic standard errors estimation. Section 3 describes the data and presents the empirical results of the paper. Section 4 concludes the paper.

2 Theoretical framework

2.1 Basic setup and average treatment effect

Current approach to program evaluation builds on the notion of potential outcomes introduced by Rubin (1974). Consider a random sample of size N from a large population. Let $T_i \in \{0, 1\}$ indicate the participation of individual i in the program and let $Y_i(0)$ denote the outcome for an individual i if he did not participate in the program and $Y_i(1)$ outcome for the individual i if he did participate in the program. Since each individual can either participate or not participate in the program, but not both, we can observe only one of these potential outcomes and the other becomes counterfactual. The response Y_i that is actually observed is then

$$Y_i = Y_i(0)(1 - T_i) + Y_i(1)T_i = \begin{cases} Y_i(0), & \text{if } T_i = 0 \\ Y_i(1), & \text{if } T_i = 1. \end{cases} \tag{1}$$

The potential outcome framework provides a definition of a causal effect, or treatment effect, for unit i as $Y_i(1) - Y_i(0)$, which is not identifiable. Particularly, under some assumptions, we can identify a population average treatment effect:

$$\tau \equiv E[Y_i(1) - Y_i(0)]. \tag{2}$$

Another frequently used estimand is the population average treatment effect on the treated:

$$\tau^{treat} \equiv E[(Y_i(1) - Y_i(0))|T_i = 1]. \tag{3}$$

To solve the identification problem following from nonobservability of both $Y_i(0)$ and $Y_i(1)$ at the same time we need to take into account the relationship between treatment assignment and the potential outcomes. In case that the assignment to treatment is randomized in such a way that it is independent of covariates as well as the potential outcomes, estimator for the average treatment effect can be obtained as a difference of means of treated and control group. However data from the observational studies can rarely be treated as randomized experiments. Comparing the average outcome of treated with the average outcome of the non-treated would not give a reasonable estimate, because groups of participants and non-participants are likely to be subject to self-selection with different outcomes even in the absence of the program participation. Therefore to estimate the causal effects an unconfoundedness is required for identification. An unconfoundedness assumption (Rosenbaum and Rubin 1983) states that the treatment is independent of the potential outcomes conditional on the observed covariates, or pretreatment variables. Formally:

Assumption 1 (Unconfounded Treatment Assignment) [2]

$$T \perp (Y(0), Y(1)) \mid X. \tag{4}$$

To satisfy this assumption, assignment to the treatment needs to be dependent only on the observed variables and we assume there are no unobservable characteristics that influence both the treatment assignment and the potential outcomes. Under unconfoundedness assumption we can estimate the average treatment effect conditioned on the covariates, $\tau(x) = E[Y(1) - Y(0)|X = x]$, by comparing a group of treated individuals and a group of non-treated individuals with the same value of covariates:

$$\begin{aligned}\tau(x) =& \mathbb{E}[Y(1) - Y(0)|X = x] \\ =& \mathbb{E}[Y(1)|T = 1, X = x] - \mathbb{E}[Y(0)|T = 0, X = x] \\ =& \mathbb{E}[Y|T = 1, X = x] - \mathbb{E}[Y|T = 0, X = x]. \end{aligned} \tag{5}$$

By averaging $\tau(x)$ over the distribution of X we then obtain the population average treatment effect:

$$\begin{aligned}\mathbb{E}[\tau(x)] =& \mathbb{E}[\mathbb{E}[Y(1) - Y(0)|X = x]] \\ =& \mathbb{E}[Y(1) - Y(0)] \qquad (6) \\ =& \mu_1 - \mu_0 = \tau, \qquad (7)\end{aligned}$$

where μ_1 and μ_0 denote unconditional means of outcomes for treated and non-treated respectively. Dimension of X can be large and instead of conditioning on all covariates, Rosenbaum and Rubin (1983) show that unconfoundedness implies that treatment and potential outcomes are also independent conditioning on the

[2] Here and in the following I omit the subscript i to simplify the notation.

propensity score $p(x) \equiv \Pr(T = 1|X = x) = \mathbb{E}[T|X = x]$, the conditional probability of receiving treatment given covariates:

$$T \perp (Y(0), Y(1)) \mid p(X). \tag{8}$$

We will assume that for all values of covariates, propensity score is bounded away from zero and one.

Assumption 2 (Overlap)

$$0 < \Pr(T = 1|X = x) < 1. \tag{9}$$

It is called an overlap assumption, as it implies that the support of the conditional distribution of X given $T = 1$ overlaps completely with that of the conditional distribution of X given $T = 0$. If there are values of covariates for which the probability of receiving the treatment is zero or one, this means that we observe only treated or control units for these values and we therefore cannot compare treated and control units at such values.

There are several methods used in the literature for estimation of average treatment effects under unconfoundedness. Among them regression methods that estimate the conditional expectations, $\mathbb{E}[Y(t)|X]$, for $t \in \{0, 1\}$, from which unconditional expectations μ_t are obtained, other use the propensity score in various ways, either by weighting treated and control outcomes or by matching treated units to control units, using the values of covariates or propensity score to match. In the paper I will particularly pay attention to the weighted methods and the combination of weighted and regression methods.

2.2 Weighting using the propensity score

To obtain an unbiased estimator of τ, we estimate μ_1 and μ_0 by reweighting of treated and control observations in the following way. Consider

$$\mathbb{E}\left[\frac{TY}{p(X)}\right] = \mathbb{E}\left[\frac{TY(1)}{p(X)}\right] \tag{10}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\frac{TY(1)}{p(X)}\middle| X\right]\right] \tag{11}$$

$$= \mathbb{E}\left[\frac{\mathbb{E}(T|X)\mathbb{E}(Y(1)|X)}{p(X)}\right] \tag{12}$$

$$= \mathbb{E}[\mathbb{E}[Y(1)|X]] \tag{13}$$

$$= \mathbb{E}[Y(1)] = \mu_1, \tag{14}$$

where the first equality follows from (1) together with $T(1-T) = 0$, second and final equality are by the low of iterated expectations, the third by unconfoundedness and fourth by $E[T|X] = p(X)$. Similarly, $\mathbb{E}\left[\frac{(1-T)Y}{1-p(X)}\right] = \mathbb{E}[Y(0)] = \mu_0$. Together it implies:

$$\mathbb{E}\left[\frac{TY}{p(X)} - \frac{(1-T)Y}{1-p(X)}\right] = \mathbb{E}[Y(1) - Y(0)] = \mu_1 - \mu_0 = \tau, \tag{15}$$

what suggests the estimator for τ that is due to Horvitz and Thompson (1952):

$$\hat{\tau}_{IPW1} = N^{-1}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - N^{-1}\sum_{i=1}^{N}\frac{(1-T_i)Y_i}{1-\hat{p}(X_i)}, \tag{16}$$

where $\hat{p}(X_i)$ denotes the estimated propensity score. To estimate the conditional probability of treatment we usually employ a parametric model, $\hat{p}(X_i) = p(X_i, \hat{\alpha})$ where α denotes the vector of parameters. Typically probit or logit models are used where the parameters of the model are estimated by maximum likelihood estimation. Given that the parametric model for the propensity score is correctly specified the estimator of average treatment effect is consistent and $\sqrt{N}$ asymptotically normally distributed. This follows by application of the generalized method of moments estimation, in this framework also often referred to as M-estimation (see Lunceford and Davidian 2004). The estimator weights the observations in each group by the inverse of the probability of being in that group, hence denoted IPW, "inverse probability weighted" estimator. By weighting the treatment and the control group we essentially obtain the unconditional expectations of response variable under treatment and non-treatment, respectively. Estimating their sample counterparts and taking the difference then directly leads to the average treatment effect estimation. If the propensity score is well-specified, the weights of the estimator should approximately add up to one. However, for finite samples for some data generating processes, the sum of the weights can depart substantially from one (see Busso, DiNardo and McCrary 2009). Therefore, another commonly used estimator in which the weights are normalized so that they add up to one in each group (see Imbens 2004) is:

$$\hat{\tau}_{IPW2} = \sum_{i=1}^{N} \frac{T_i Y_i}{\hat{p}(X_i)} \Bigg/ \sum_{i=1}^{N} \frac{T_i}{\hat{p}(X_i)} - \sum_{i=1}^{N} \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} \Bigg/ \sum_{i=1}^{N} \frac{1-T_i}{1-\hat{p}(X_i)}. \tag{17}$$

Note that $\mathbb{E}\left[\frac{T}{p(X)}\right] = \mathbb{E}\left[\frac{\mathbb{E}(T|X)}{p(X)}\right] = 1$ and $\mathbb{E}\left[\frac{1-T}{1-p(X)}\right] = 1$ so the term in (17) estimates the difference of the unconditional expectations as in the previous case furthermore let the weights add up to one. However, as shown in Robins and Rotnitzky (1995), and Hahn (1998), even if we know the true propensity score, IPW1 and IPW2 lead in general to inefficient estimates. In fact, it is better, in terms of large sample efficiency, to weight using the estimated rather than the true propensity score. Rubin and Thomas (1996) conclude that using the estimated rather than known propensity score can improve the efficiency of the estimator when the propensity score belongs to a parametric family. Hirano, Imbens and Ridder (2003) show that in a special case, when the propensity score is estimated nonparametrically, the resulting estimate is asymptotically efficient.

Lunceford and Davidian (2004) provide another estimator that is an asymptotically optimal variance minimizing combination of the former two estimators, IPW1 and IPW2:

$$\begin{aligned}\hat{\tau}_{IPW3} = & \sum_{i=1}^{N} \frac{T_i Y_i}{\hat{p}(X_i)} \left(1 - \frac{C_1}{\hat{p}(X_i)}\right) \Bigg/ \left\{ \sum_{i=1}^{N} \frac{T_i}{\hat{p}(X_i)} \left(1 - \frac{C_1}{\hat{p}(X_i)}\right) \right\} \\ & - \sum_{i=1}^{N} \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} \left(1 - \frac{C_0}{\hat{p}(X_i)}\right) \Bigg/ \left\{ \sum_{i=1}^{N} \frac{1-T_i}{1-\hat{p}(X_i)} \left(1 - \frac{C_0}{\hat{p}(X_i)}\right) \right\},\end{aligned} \tag{18}$$

where $C_1 = \sum_{i=1}^{N} \frac{T_i - \hat{p}(X_i)}{\hat{p}(X_i)} \,/\, \sum_{i=1}^{N} \left\{\frac{T_i - \hat{p}(X_i)}{\hat{p}(X_i)}\right\}^2$ and $C_0 = \sum_{i=1}^{N} \frac{T_i - \hat{p}(X_i)}{1-\hat{p}(X_i)} \,/\, \sum_{i=1}^{N} \left\{\frac{T_i - \hat{p}(X_i)}{1-\hat{p}(X_i)}\right\}^2$.

It is derived using the equalities

$$\sum_{i=1}^{N}\left\{\frac{T_i(Y_i-\mu_1)}{p(X_i)}+\eta_1\left(\frac{T_i-p(X_i)}{p(X_i)}\right)\right\}=0, \tag{19}$$

$$\sum_{i=1}^{N}\left\{\frac{(1-T_i)(Y_i-\mu_0)}{1-p(X_i)}-\eta_0\left(\frac{T_i-p(X_i)}{1-p(X_i)}\right)\right\}=0, \tag{20}$$

that both $\hat{\tau}_{IPW1}$ and $\hat{\tau}_{IPW2}$ solve with $(\eta_0,\eta_1)=(\mu_0,\mu_1)$ and $(\eta_0,\eta_1)=(0,0)$, respectively. Identifying η_0 and η_1 minimizes the large sample variance of solutions to the equations (19) and (20) and improves upon $\hat{\tau}_{IPW1}$ and $\hat{\tau}_{IPW2}$. They are estimated solving

$$\sum_{i=1}^{N}\left\{\frac{T_i(Y_i-\mu_1)}{p(X_i)^2}+\eta_1\left(\frac{T_i-p(X_i)}{p(X_i)}\right)^2\right\}=0, \tag{21}$$

$$\sum_{i=1}^{N}\left\{\frac{(1-T_i)(Y_i-\mu_0)}{(1-p(X_i))^2}-\eta_0\left(\frac{T_i-p(X_i)}{1-p(X_i)}\right)^2\right\}=0. \tag{22}$$

Solving these equations the estimator in (18) can be obtained. Lunceford and Davidian (2004) argue that particularly for small N, inverse weighting of observation by a very small value of propensity score can result in numerical instability, what is in IPW3 estimator taken care of by the proper scaling of each weight.

2.3 Doubly-robust estimator

Compared to weighting methods that estimate the average treatment effects by estimating the propensity score $p(X)$ and use it to weight the treated and control outcomes, regression methods are based on estimating the two conditional means $\mu_1(x)=\mathbb{E}[Y(1)|X=x]$ and $\mu_0(x)=\mathbb{E}[Y(0)|X=x]$. As shown in (5),

$$\mu_1(x)=\mathbb{E}[Y(1)|X=x]=\mathbb{E}[Y|T=1,X=x], \tag{23}$$

$$\mu_0(x)=\mathbb{E}[Y(0)|X=x]=\mathbb{E}[Y|T=0,X=x], \tag{24}$$

and we can estimate $\mu_0(.)$ using regression methods for the untreated subsample and $\mu_1(.)$ for the treated subsample. Given consistent estimators $\hat{\mu}_1(x)$ and $\hat{\mu}_0(x)$ we obtain a consistent estimator for unconditional means μ_1, μ_0 and their difference, average treatment effect $\tau=\mu_1-\mu_0$, by averaging them

$$\hat{\tau}_{REG}=N^{-1}\left(\sum_{i=1}^{N}\hat{\mu}_1(X_i)-\sum_{i=1}^{N}\hat{\mu}_0(X_i)\right). \tag{25}$$

To estimate $\mu_1(x)$ and $\mu_0(x)$ we can specify the parametric models $m_1(X, \beta_1)$ and $m_0(X, \beta_0)$ which in the simplest case can be linear functions of parameters:

$$\mu_1(x) = m_1(X, \beta_1) = X\beta_1,$$
$$\mu_0(x) = m_0(X, \beta_0) = X\beta_0,$$

where $X \in \mathbb{R}^{N\times(k+1)}$ includes k covariates and the vector of ones and β_t for $t \in \{0, 1\}$ is the k-dimensional vector. This model is relevant in case the outcome variable takes continuous values. Alternatively X can be replaced by any other function of X to describe the structure of the outcome data better, or we can use a different model for example a logit model which is more appropriate in case of the binary outcome variable. We can also go beyond the parametric models and use nonparametric methods when modeling the propensity score. For the discussion on regression methods used in the literature see Imbens and Wooldridge (2008).

For both methods, weighting and regression, the estimation greatly depends on the parametric models that we choose to approximate the means of the potential outcomes or the propensity score, and the estimators can suffer from inconsistency in case of the model misspecification. Very desirable in these cases would be an approach that uses a combination of regression and the propensity score weighting method to achieve some robustness against model misspecification. This approach is adopted by Robins, Rotnitzky and Zhao (1995) who develop the so-called doubly-robust estimator[3] relying on the combination of the two methods:

$$\begin{aligned}\hat{\tau}_{DR} =& N^{-1}\sum_{i=1}^{N}\left(\frac{T_iY_i}{\hat{p}(X_i)} - \left(\frac{T_i}{\hat{p}(X_i)} - 1\right)m_1(X_i, \hat{\beta}_1)\right)\\ &- N^{-1}\sum_{i=1}^{N}\left(\frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} - \left(\frac{1-T_i}{1-\hat{p}(X_i)} - 1\right)m_0(X_i, \hat{\beta}_0)\right), \end{aligned} \quad (26)$$

where $\hat{\beta}_t$, for $t \in \{0, 1\}$ comes from the regression model estimation of conditional means $\mu_t(x)$ by models $m_t(X, \beta_t)$. Each term in $\hat{\tau}_{DR}$ has the form of the weighted estimator but "augmented" by the expression involving the regression what increases the efficiency of the weighted estimator. But it is particularly a so-called "double-robustness" property that motivates this estimator, which says

[3]There are several ways of combining the two estimation approaches, I focus only on this specific one.

that the estimator remains consistent if either the propensity score model is correctly specified but the two regression models m_0 and m_1 are not or the two regression models are correctly specified but the propensity score model is not and is automatically consistent for the correct specification of all models. Consistency of the estimator for the two cases of misspecification is shown in Appendix. Nevertheless, in case of misspecification of one of the two models the estimator need no longer be more efficient than the weighted estimator (Lunceford and Davidian 2004). But neither weighted estimators need be consistent if $p(X)$ is incorrectly specified. Also in case of the correctly specified regression models the variance of the estimator may be larger in large samples compared to the regression estimator, it only offers the protection in case of misspecification (Davidian 2007).

2.4 Asymptotic properties of estimators

Large sample properties of weighted and doubly-robust estimators can be obtained directly from the theory of generalized method of moments estimation. Using the notation of Wooldridge (2001) let $W_i \in \mathbb{R}^M;\ i = 1, ..., N$ be a vector of observations and θ a P-dimensional vector of parameters. Assume that for a known function $\psi(W_i, \theta) \in \mathbb{R}^L$, the parameter $\theta_0 \in \Theta \subset \mathbb{R}^P$ satisfies the moment conditions

$$\mathbb{E}[\psi(W_i, \theta_0)] = 0. \tag{27}$$

To identify θ_0, moment conditions must satisfy $L \geq P$. In a simple case when $L = P$, a method of moments estimator, $\hat{\theta}$, solves the sample counterpart of (27),

$$\frac{1}{N}\sum_{i=1}^{N} \psi(W_i, \hat{\theta}) = 0. \tag{28}$$

If there is a unique θ_0 for which $\mathbb{E}[\psi(W_i, \theta_0)] = 0$, then by the weak law of large numbers $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \rightarrow \infty$. Under standard regularity conditions, the estimator is asymptotically normally distributed,

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, A(\theta_0)^{-1}B(\theta_0)(A(\theta_0)^{-1})'), \tag{29}$$

where $A(\theta_0) = \mathbb{E}\left[-\frac{\partial\psi(W_i, \theta_0)}{\partial\theta_0'}\right]$ and $B(\theta_0) = \mathbb{E}[\psi(W_i, \theta_0)\psi(W_i, \theta_0)']$.

In this framework, we can specify the vectors of parameters θ and functions $\psi(W_i, \theta)$ for each estimator of τ_{IPW1}, τ_{IPW2}, τ_{IPW3} and τ_{DR} that become zero in expectation and thus satisfy the moment conditions. General properties of method of moments estimator then imply the consistency of the estimators and the asymptotic variance of the estimators can be derived in the following way[4].

First weighting estimator $\hat{\tau}_{IPW1}$ involves estimating $\hat{\mu}_1$, $\hat{\mu}_0$ and $\hat{\alpha}$, the parameter vector of estimated propensity score. In the following I assume a probit estimation of propensity score $p(X) = p(X, \alpha) = \Phi(X\alpha)$, where Φ denotes the cumulative distribution function of the standard normal distribution. The estimated parameter vector $\hat{\theta}_{IPW1} = (\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_0)$ then solves the following moment condition

$$\frac{1}{N}\sum_{i=1}^{N}\psi(W_i, \hat{\theta}_{IPW1}) = \frac{1}{N}\sum_{i=1}^{N}\psi((Y_i, X_i), (\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_0)) = 0, \tag{30}$$

where

$$\psi((Y_i, X_i), (\alpha, \mu_1, \mu_0)) = \begin{pmatrix} \psi_1((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_2((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_3((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \end{pmatrix} = \begin{pmatrix} \frac{T_i - \Phi(X_i'\alpha)}{\Phi(X_i'\alpha)(1-\Phi(X_i'\alpha))}\frac{\partial \Phi(X_i'\alpha)}{\partial \alpha} \\ \frac{T_i Y_i}{\Phi(X_i'\alpha)} - \mu_1 \\ \frac{(1-T_i)Y_i}{1-\Phi(X_i'\alpha)} - \mu_0 \end{pmatrix}. \tag{31}$$

The first moment condition, $\mathbb{E}[\psi_1] = 0$, follows from the maximum likelihood estimation of α and is a score for a probit model. The other two equations come from the weighting identification of μ_1 and μ_0. Applying (29) we have

$$\sqrt{N}(\hat{\theta}_{IPW1} - \theta_0) \xrightarrow{d} N(0, A_{IPW1}^{-1}B_{IPW1}(A_{IPW1}^{-1})') = N(0, AVar_{IPW1}), \tag{32}$$

where

$$A_{IPW1} = -\mathbb{E}\begin{bmatrix} \frac{\partial \psi_1(W_i, \theta_{IPW1})}{\partial \alpha'} & \frac{\partial \psi_1(W_i, \theta_{IPW1})}{\partial \mu_1} & \frac{\partial \psi_1(W_i, \theta_{IPW1})}{\partial \mu_0} \\ \frac{\partial \psi_2(W_i, \theta_{IPW1})}{\partial \alpha'} & \frac{\partial \psi_2(W_i, \theta_{IPW1})}{\partial \mu_1} & \frac{\partial \psi_2(W_i, \theta_{IPW1})}{\partial \mu_0} \\ \frac{\partial \psi_3(W_i, \theta_{IPW1})}{\partial \alpha'} & \frac{\partial \psi_3(W_i, \theta_{IPW1})}{\partial \mu_1} & \frac{\partial \psi_3(W_i, \theta_{IPW1})}{\partial \mu_0} \end{bmatrix} \tag{33}$$

[4] See Uysal (2011) and Lunceford and Davidian (2004) for further reference.

and

$$B_{IPW1} = \mathbb{E}\left[\begin{pmatrix} \psi_1(W_i, \theta_{IPW1}) \\ \psi_2(W_i, \theta_{IPW1}) \\ \psi_3(W_i, \theta_{IPW1}) \end{pmatrix} \begin{pmatrix} \psi_1(W_i, \theta_{IPW1})' & \psi_2(W_i, \theta_{IPW1})' & \psi_3(W_i, \theta_{IPW1})' \end{pmatrix} \right]. \tag{34}$$

The asymptotic distribution of $\hat{\tau}_{IPW1}$ is then

$$\sqrt{N}(\hat{\tau}_{IPW1} - \tau_0) \xrightarrow{d} N(0, AVar_{IPW1}(\hat{\mu}_1) + AVar_{IPW1}(\hat{\mu}_0) - 2ACov_{IPW1}(\hat{\mu}_1, \hat{\mu}_0)), \tag{35}$$

where the asymptotic variance of $\hat{\tau}_{IPW1}$ uses the 2×2 submatrix of the $AVar_{IPW1}$ matrix that carry an information on the asymptotic variances and covariance of $\hat{\mu}_1$ and $\hat{\mu}_0$ estimators. This finally allows to compute the asymptotic standard errors of $\hat{\tau}_{IPW1}$. Formula (35) applies to all the remaining estimators and I therefore list only the relevant moment conditions for joint vector of parameters in each case.

For the estimator $\hat{\tau}_{IPW2}$ estimator the following moment conditions hold:

$$\frac{1}{N}\sum_{i=1}^{N} \psi(W_i, \hat{\theta}_{IPW2}) = \frac{1}{N}\sum_{i=1}^{N} \psi((Y_i, X_i), (\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_0)) = 0, \tag{36}$$

where

$$\psi((Y_i, X_i), (\alpha, \mu_1, \mu_0)) = \begin{pmatrix} \psi_1((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_2((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_3((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \end{pmatrix} = \begin{pmatrix} \frac{T_i - \Phi(X_i'\alpha)}{\Phi(X_i'\alpha)(1-\Phi(X_i'\alpha))} \frac{\partial \Phi(X_i'\alpha)}{\partial \alpha} \\ \frac{T_i(Y_i - \mu_1)}{\Phi(X_i'\alpha)} \\ \frac{(1-T_i)(Y_i - \mu_0)}{1-\Phi(X_i'\alpha)} \end{pmatrix}. \tag{37}$$

The third weighted estimator $\hat{\tau}_{IPW3}$ can be estimated using the moment conditions:

$$\frac{1}{N}\sum_{i=1}^{N} \psi(W_i, \hat{\theta}_{IPW3}) = \frac{1}{N}\sum_{i=1}^{N} \psi((Y_i, X_i), (\hat{\alpha}, \hat{\mu}_1, \hat{\mu}_0)) = 0, \tag{38}$$

$$\psi((Y_i, X_i), (\alpha, \mu_1, \mu_0)) = \begin{pmatrix} \psi_1((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_2((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \\ \psi_3((Y_i, X_i), (\alpha, \mu_1, \mu_0)) \end{pmatrix} = \begin{pmatrix} \frac{T_i - \Phi(X_i'\alpha)}{\Phi(X_i'\alpha)(1-\Phi(X_i'\alpha))} \frac{\partial \Phi(X_i'\alpha)}{\partial \alpha} \\ \frac{T_i(Y_i - \mu_1)}{\Phi(X_i'\alpha)} + \eta_1 \frac{T_i(Y_i - \Phi(X_i'\alpha))}{\Phi(X_i'\alpha)} \\ \frac{(1-T_i)(Y_i - \mu_0)}{1-\Phi(X_i'\alpha)} - \eta_0 \frac{T_i(Y_i - \Phi(X_i'\alpha))}{1-\Phi(X_i'\alpha)} \end{pmatrix}, \tag{39}$$

where η_1 and η_0 are estimated by solving (21) and (22).

The method of moments estimation of the doubly-robust estimator includes two additional moment conditions related to the estimation of coefficients β_1 and β_0 of the regression models for conditional means.

$$\frac{1}{N}\sum_{i=1}^{N}\psi(W_i,\hat{\theta}_{DR}) = \frac{1}{N}\sum_{i=1}^{N}\psi((Y_i,X_i),(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_0,\hat{\mu}_1,\hat{\mu}_0)) = 0 \tag{40}$$

$$\psi((Y_i,X_i),(\alpha,\beta_1,\beta_0,\mu_1,\mu_0)) = \begin{pmatrix} \frac{T_i-\Phi(X_i'\alpha)}{\Phi(X_i'\alpha)(1-\Phi(X_i'\alpha))}\frac{\partial\Phi(X_i'\alpha)}{\partial\alpha} \\ T_iX_i(Y_i-X_i'\beta_1) \\ (1-T_i)X_i(Y_i-X_i'\beta_0) \\ \frac{T_iY_i}{\Phi(X_i'\alpha)} - \left(\frac{T_i}{\Phi(X_i'\alpha)}-1\right)m_1(X_i,\beta_1) \\ \frac{(1-T_i)Y_i}{1-\Phi(X_i'\alpha)} - \left(\frac{1-T_i}{1-\Phi(X_i'\alpha)}-1\right)m_0(X_i,\beta_0) \end{pmatrix}, \tag{41}$$

where the second and the third moment conditions come from the minimization problem related to the regressions m_1 and m_0, in case of the linear regression models it is the minimization of the sum of squared residuals.

3 Empirical analysis

3.1 Data description

In the paper I estimate an impact of active labour market programs in Austria using an administrative data set from a study of Hofer, Sellner and Weber (2007). Data set was obtained at the Applied Research Department of the Institute for Advanced Studies, Vienna. Data combines an information from the social security records and from the registers of the Austrian public employment office (AMS). This data set includes individuals who entered unemployment from March to August 1999. The spring-summer season was chosen to avoid differences caused by a seasonal unemployment. For the selected sample, authors collected all active labour market program spells during the period 1997 to 2004 and various relevant personal characteristics of the individuals. In the paper I focus on two particular types of programs, active job-search programs and training programs and evaluate their impact on the change in employment measured by days of employment in four years after the program start. Participants of either active job-search programs or training programs form a treatment group, unemployed not participating in any active labour market program form a control group, whereas participants of unclassified and other programs are excluded. Also participants younger than 20 and older than 50, unemployed less than 15 days when their first program spell started and those who participated in the program after 2000 are dropped from a sample. Table 1 shows the distribution of respective program types by gender.

Program type	**Male**	**Female**	**Total**
Active job-search programs	3 339	3 583	6 922
Training programs	1 904	2 913	4 817
Non-participants	23 275	26 883	50 158
Total	28 518	33 379	61 897

Table 1: Number of observations by program type and gender

In order to construct an outcome variable for the control group, number of employment days for non-participants was calculated from hypothetical program starts generated according to the procedure suggested by Lechner et al. (2004) and for this dataset described in details in Hofer, Sellner and Weber (2007).

3.2 Empirical results

I evaluate the average treatment effect active labour market programs have on the post-program employment using methods based on propensity score weighting as described in the previous section. In the initial study, Hofer, Sellner and Weber (2007) estimate the average treatment effect on treated for two program types, active job-search programs and training programs using propensity score matching approach. They find, that in the medium run (4 years), active job-search programs are more effective than the traditional training programs. However, for both types of programs, a significant effect on post-employment was found only for females and amount to 85.7 days for active job-search and 62 days for training programs.

In their analysis they matched participants and non-participants on the propensity score basis applying nearest neighbor matching. I use an alternative approach in this paper and estimate the average treatment effects based on the propensity score weighting methods and doubly-robust approach which combines weighting with the regression. All these estimators rely on the propensity score estimation. I use a probit model to estimate the propensity score. Explanatory variables used to model the probability of receiving a treatment are pre-treatment variables such as personal characteristics, education, last occupation, geography and labour market history. I include the regional dummies to account for regional labour market situation, also dummies for month of entry into unemployment to control for seasonality and other interaction dummies. Furthermore, data for treatment group and control group come from the same dataset measured at the same point of time. This rich dataset then makes the unconfoundedness assumption more plausible. To justify the overlap assumption visual inspection of propensity score distribution is used. For each program type for females and males separate histograms are plotted for treated and control group to make sure there is a sufficient overlap in the support of the conditional distribution between the two groups. Within all subgroups the overlap seems satisfactory, nevertheless for some parts of the distribution we may observe the cases where there is a very low density especially for the values of the propensity score above 0.7. The overlap is however important for the estimation because weighting by the values that are close to zero in denominator for some units may cause difficulties to obtain the precise estimate for treatment effect (Imbens and Wooldridge 2008, section 5.4). Several methods have been suggested to determine the common support region. They involve dropping the units that have no counterpart in the other group what guarantees we will estimate only on the region with the common support where both treated and control units are available. I apply the Minima and Maxima comparison criterion as described in Caliendo and Kopeinig (2005) and drop the individuals whose propensity score is above the highest propensity score value of the control group and below the lowest value of the treated group. The resulting histograms are shown in Figures 1- 4. When reading the graphs it is important to notice that the scale for the treated and for the control group differs significantly as the proportion of untreated individuals is much higher in all subsamples. All the covariates used in the specification of propensity score are listed in the Table 4. Suggested set of variables coincides with the one in the original paper. Probit estimation results are reported in the Table 5.

Given the estimated propensity score I first use three weighted estimators to estimate the average treatment effects of two types of programs separately for men and women. The first three columns of Table 2 report the results for each of the four groups considered. There is a positive treatment effect for both females and males participating in active job-search programs, but only the first estimates of τ_{IPW1} are significant on the 5% significance level in both cases. It suggests that the treatment effect of the program, in terms of days employed four years after the program entry, is 98.72 days and 82.33 days for females and males respectively. However, other estimates considered in the literature more often, IPW2 and IPW3, differ from the first estimate in both cases and are not significant. On the other hand, for male participants, there is an unambiguously

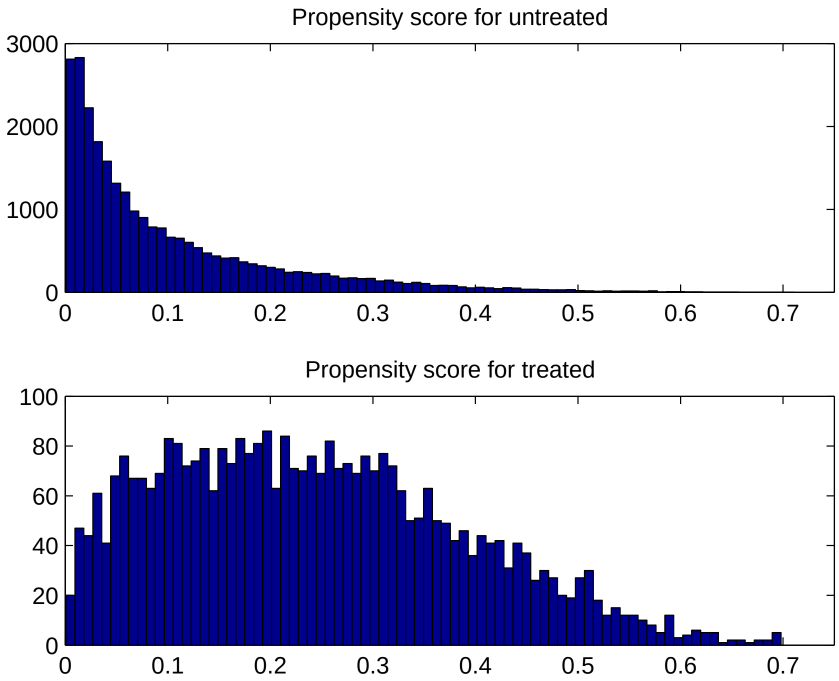

Figure 1: Histograms of propensity scores estimated for Female - Active job-search programs

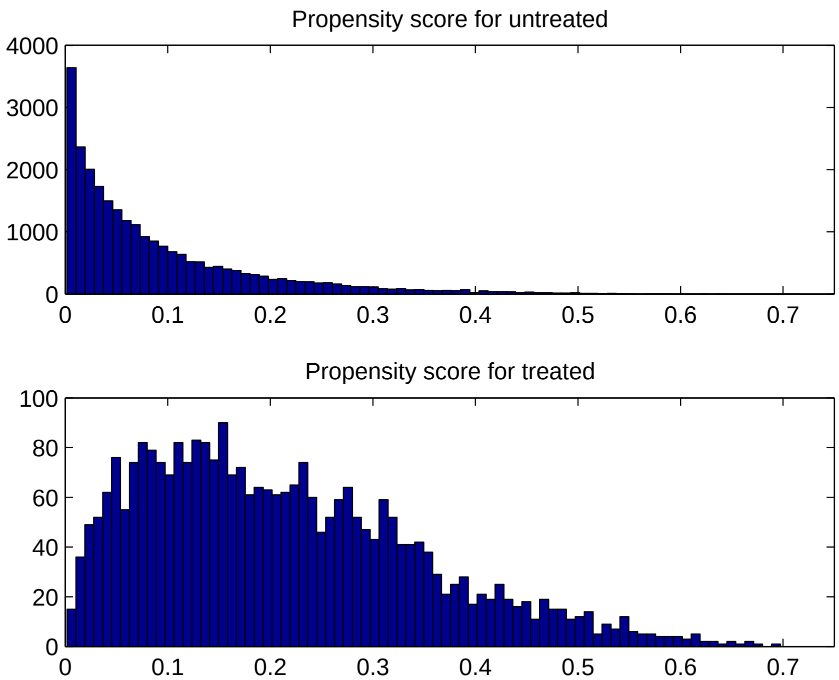

Figure 2: Histograms of propensity scores estimated for Female - Training programs

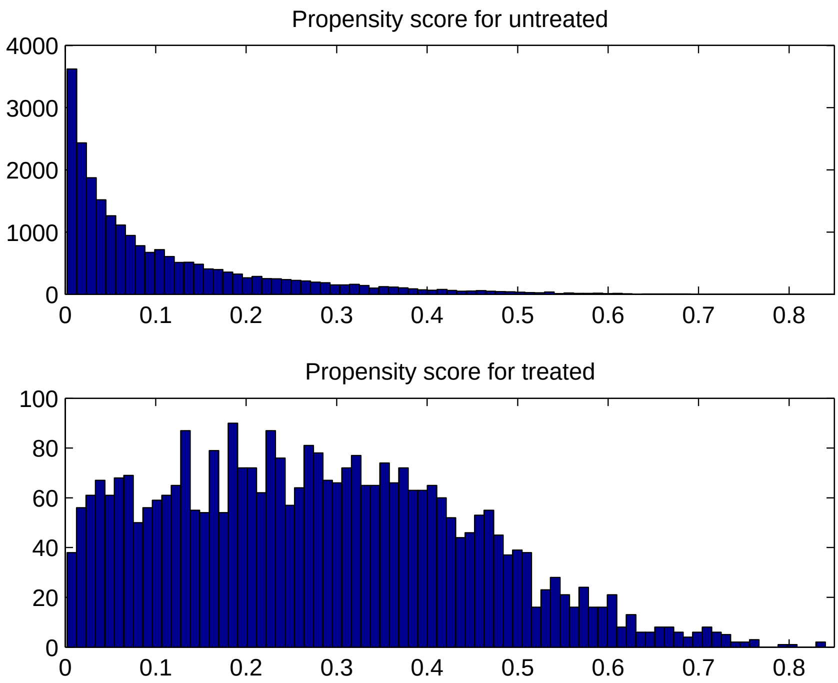

Figure 3: Histograms of propensity scores estimated for Male - Active job-search programs

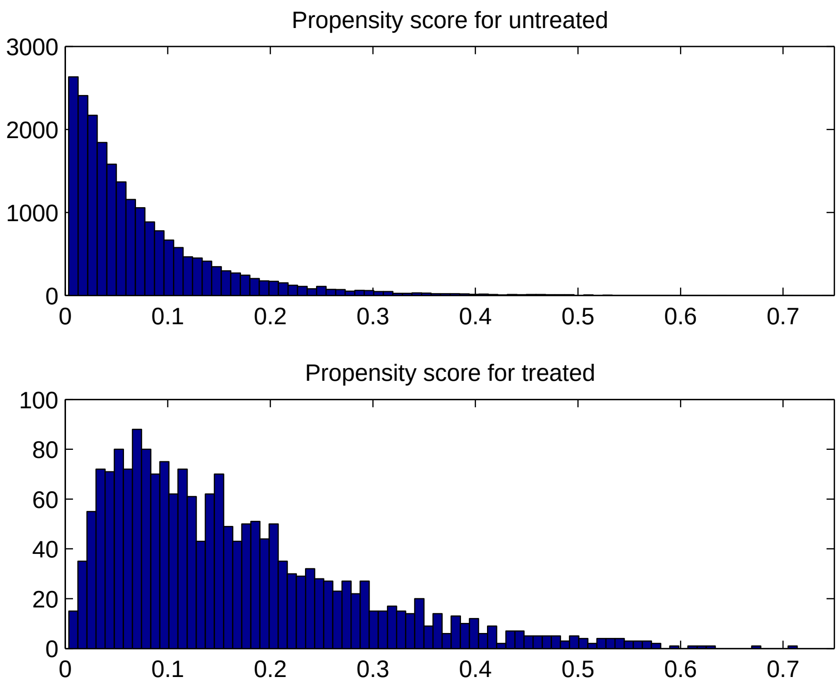

Figure 4: Histograms of propensity scores estimated for Male - Training programs

negative treatment effect estimated for group of training program participants. For IPW1, IPW2 and IPW3 estimators it ranges from -58.36 to -68.08 and is significant in all three cases.

	$\hat{\tau}_{\mathbf{IPW1}}$	$\hat{\tau}_{\mathbf{IPW2}}$	$\hat{\tau}_{\mathbf{IPW3}}$	$\hat{\tau}_{\mathbf{DR}}$
Female				
Active job search programs	98.718**	35.549*	25.439	50.887**
	(48.886)	(20.463)	(21.891)	(24.389)
Training programs	27.211	26.230	26.193	32.621**
	(30.424)	(17.861)	(17.866)	(16.412)
Male				
Active job search programs	82.333**	-17.205	-22.594	-13.085
	(36.625)	(17.738)	(20.776)	(18.598)
Training programs	-68.077***	-58.355***	-58.836***	-37.895***
	(19.931)	(16.804)	(17.257)	(13.544)

Table 2: Estimated average treatment effects. Asymptotic standard errors in parentheses. *, ** and *** denote significance at 10%, 5% and 1% significance level.

Next I estimate the treatment effect of the programs using doubly-robust estimator which provides some robustness against possible misspecification of parametric models for either the conditional mean in the regression part or the propensity score for weighting. Results are reported in the last column of Table 2. It estimates positive significant treatment effects of both active job-search and training programs for females of the size 50.89 days in the first case and 32.62 days in the second case. Significant is also a negative effect of training programs for males that amounts to 37.90 days decrease in employment. From the results we see that in terms of the size, the estimates indicate that active job-search programs are followed by a larger increase in employment (or milder decrease of employment) compared to training programs for both men and women. This may support the argument, that participants of training programs are subject to stronger lock-in effect. Compared to active job-search programs they cannot search for a job during the program. In the short run a negative treatment effect can therefore occur, as we observe in case of males, and is expected to be positive rather in the long run. Also the difference in effects between men and women can be found as it is documented in the similar studies and can reflect the fact that the higher increase in the female employment may be a result of the reduction of share of women leaving the labour force due to pregnancy.

For comparison I list the results of average treatment effects in all subgroups estimated by the nearest neighbor matching in the Appendix. Significant treatment effect is reported only for the group of male participants in the training programs that amounts to 64.19 days decrease in employment. Results of other groups are, compared with the doubly-robust estimator, smaller but all are not significant. Matching estimators assign for the missing potential outcomes the outcomes of a few (in this case one) nearest neighbors of the opposite treatment group where matching is commonly done on the values of propensity score. Average treatment effect is then computed as an average of the outcomes for the matches. This estimator can however be biased due to possible discrepancies between the covariates of the matched observations and their matches (Imbens and Wooldridge 2008, section 5.5). Abadie and Imbens (2006) also show that matching estimators are generally not efficient and even in the case where the bias is of low enough order to be dominated by the variance, the estimators do not reach the efficiency bound given a fixed number of matches.

Finally I provide also the basic difference between the treatment and the control group (Table 3), that does not adjust for any covariates, expressed by the difference in the average outcomes in the two groups. We see that outcomes in groups of treated versus nontreated individuals differ and except for the group of females in the training programs, difference is negative. This would suggest, that by ignoring the dependence of potential outcomes and treatment assignment

Treatment group	**Female**	**Male**
Active job-search programs	-6.273	-89.157
Training programs	45.230	-86.224

Table 3: Mean difference of outcomes between the treated and the control group

on the covariates, particularly in three treatment groups is outcome, number of employment days for treated, lower than for the group of not treated.

To inspect the effect of the program participation on the future employment over time, for each subgroup the average treatment effect is estimated in each of 20 quarters following the program entry. For this analysis, I use the last estimator due to its favorable double robustness property. The results are depicted in Figure 5. The analysis supports the argument that the lock-in effect is stronger in case of the training programs and we can observe that a positive treatment effect occurs much later especially in case of men where the positive effect of the training programs is estimated for the first time in the eighth quarter after the program start. On the other hand very prompt reaction for females in active job-search programs is reported who experience the positive treatment effect already in the second quarter following the entry into the program.

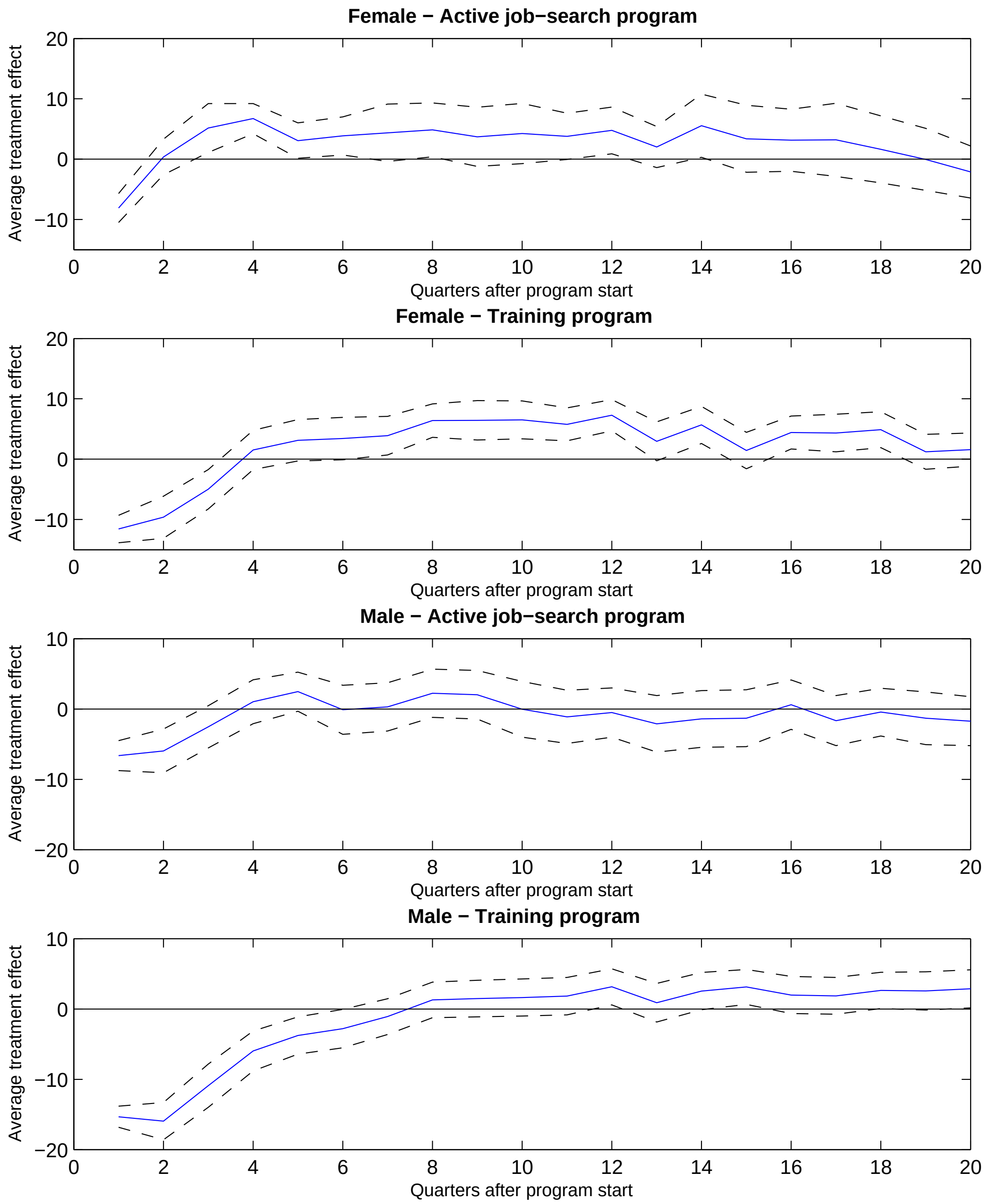

Figure 5: Average treatment effects over time. Estimated using doubly-robust method. Asymptotic standard errors used.

4 Conclusion

In the paper, I estimate the impact of two types of active labour market programs on the future employment using propensity score weighting estimators and the estimator that combines weighting and regression for estimating average treatment effect under unconfoundedness assumption. I consider three weighting estimators as proposed in Lunceford and Davidian (2004) and one doubly-robust estimator that combines regression with the weighting approach and has a useful property that the estimated parameter stays consistent even in cases when either estimated model for conditional mean functions in the regression or the model for propensity score used for weighting are misspecified. In a theoretical part I define the estimators and show their asymptotic distribution derivation using method of moments estimation what leads to the asymptotic standard errors estimation.

I then apply the estimation procedure to evaluate the effects of two types of labour market programs, active job-search programs and training programs, in terms of the future employment following four years after the program start. I use an informative individual dataset from the study of Hofer, Sellner and Weber (2007) on unemployed individuals, their personal characteristics, labour market history and their potential participation on labour market programs. I evaluate the effect of participation in the two types of programs for females and males. I find a better performance of active job-search programs compared to training programs considering the future employment increase for females. Both active job-search programs and training programs have significantly positive treatment effects for females of 50.89 days and 32.62 days, respectively, when estimated using doubly-robust estimator. In case of male participants I find the significant negative effect of training programs participation. This result is obtained using all three weighted estimators, and ranges from -58.36 to -68.08, and also by doubly-robust estimator that estimates a decrease of 37.90 days in employment after program participation. Due to the fact, that I analyse the medium run employment effect of the programs, this might be the result of stronger lock-in effect traditional training programs have compared to the active job search programs. Further evidence can be found in the additional analysis of quarterly treatment effects estimated for the time period up to 20 quarters after the program start. For training programs positive treatment effect is reported only after eighth period for men and fourth period for women whereas job-search programs report it in earlier periods, fourth period for men and second period for women.

References

Abadie, Alberto and Guido W. Imbens, "On the Failure of the Bootstrap for Matching Estimators," NBER Technical Working Papers 0325, National Bureau of Economic Research, Inc June 2006.

Busso, Matias, John DiNardo, and Justin McCrary, "Finite sample properties of semiparametric estimators of average treatment effects," forthcoming in the Journal of Business and Economic Statistics, 2009.

Caliendo, Marco and Sabine Kopeinig, "Some Practical Guidance for the Implementation of Propensity Score Matching," IZA Discussion Papers 1588, Institute for the Study of Labor (IZA), 2005.

Davidian, Marie, "Double Robustness in the Estimation of Causal Treatment Effects," Online, 2007. http://www4.stat.ncsu.edu/d̃avidian/double.pdf - accessed on: 8 June 2013.

Hahn, Jinyong, "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," Econometrica, 1998, 66 (2), 315-332.

Hirano, Keisuke, Guido Imbens, and Geert Ridder, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," Econometrica, 2003, 71, 1161-1189.

Hofer, Helmut and Andrea Weber, "Active Labor Market Policy in Austria: Practice and Evaluation Results," Vierteljahrshefte zur Wirtschaftsforschung, 2006.

Hofer, Helmut, Richard Sellner, and AndreaWeber, "Evaluating job search programmes in Austria - further evidence," Technical Report, Institute for Advanced Studies, Vienna 2007.

Horvitz, D. G. and D. J. Thompson, ""A generalisation of sampling without replacement from a finite universe," Journal of the American Statistical Association, 1952, 47, 663-685.

Imbens, Guido, "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," Review of Economics and Statistics, 2004.

Imbens, Guido M. and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," Working Paper 14251, National Bureau of Economic Research August 2008.

Kluve, Jochen, "The Effectiveness of European Active Labor Market Policy," IZA Discussion Papers 2018, Institute for the Study of Labor (IZA), 2006.

Lechner, Michael, Ruth Miquel, Conny Wunsch, "Long-Run Effects of Public Sector Sponsored Training in West Germany," IZA Discussion Papers 1443, Institute for the Study of Labor (IZA), 2004.

Lechner, Michael and Stephan Wiehler, "Kids or courses? Gender differences in the effects of active labor market policies," Journal of Population Economics, July 2011, 24 (3), 783-812.

Leuven, Edwin and Barbara Sianesi, "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing," Statistical Software Components, Boston College Department of Economics April 2003.

Lunceford, Jared K. and Marie Davidian, "Stratification and Weighting Via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," Statistics in Medicine, 2004.

Robins, J M, A Rotnitzky, and L P Zhao, "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data," Journal of the American Statistical Association, 1995, 90, 106-129.

Rosenbaum, Paul R. and Donald B. Rubin, "The central role of the propensity score in observational studies for causal effects," Biometrika, 1983, 70 (1), 41-55.

Rubin, D. B. and N. Thomas, "Matching Using Estimated Propensity Scores: Relating Theory to Practice," Biometrics, 1996, 52, 249-64.

Rubin, D.B., "Estimating causal effects of treatments in randomized and non-randomized studies," Journal of Educational Psychology, 1974, 66 (5), 688-701.

Uysal, Selver Derya, "Three Essays on Doubly Robust Estimation Methods." PhD dissertation, University of Konstanz, Konstanz, Germany 2011.

Weber, Andrea and Helmut Hofer, "Are Job Search Programs a Promising Tool? A Microeconometric Evaluation for Austria," IZA Discussion Papers 1075, Institute for the Study of Labor (IZA) March 2004.

Wooldridge, Jeffrey M., "Econometric Analysis of Cross Section and Panel Data," Vol. 1 of MIT Press Books, The MIT Press, 2001.

Appendix

A Consistency of doubly-robust estimator

By the law of large numbers, estimator $\hat{\tau}_{DR} = \hat{\mu}_{1,DR} - \hat{\mu}_{0,DR}$ defined in (26) estimates the following mean:

$$\begin{aligned}\tau_{DR} =& \mathbb{E}\left[\frac{TY}{p(X,\alpha)} - \left(\frac{T}{p(X,\alpha)} - 1\right) m_1(X,\beta_1)\right] \\ &- \mathbb{E}\left[\frac{(1-T)Y}{1-p(X,\alpha)} - \left(\frac{1-T}{1-p(X,\alpha)} - 1\right) m_0(X,\beta_0)\right], \end{aligned} \tag{42}$$

where $p(X,\hat{\alpha})$, $m_1(X,\hat{\beta}_1)$ and $m_0(X,\hat{\beta}_0)$ are replaced by the terms they estimate. Then $\hat{\mu}_{1,DR}$ estimates

$$\begin{aligned}&\mathbb{E}\left[\frac{TY(1)}{p(X,\alpha)} - \left(\frac{T}{p(X,\alpha)} - 1\right) m_1(X,\beta_1)\right] \\ &= \mathbb{E}\left[Y(1) + \left(\frac{T}{p(X,\alpha)} - 1\right)(Y(1) - m_1(X,\beta_1))\right] \\ &= \mathbb{E}[Y(1)] + \mathbb{E}\left[\left(\frac{T}{p(X,\alpha)} - 1\right)(Y(1) - m_1(X,\beta_1))\right]. \end{aligned} \tag{43}$$

For $\hat{\mu}_{1,DR}$ to estimate $\mathbb{E}[Y(1)]$, the second term in (43) must be zero.

Specified models $p(X,\alpha)$ and $m_1(X,\beta_1)$ may or may not equal to the true propensity score and true conditional expectation. Consider the following two situations in which one of the two models is misspecified:

Situation 1: Model $p(X,\alpha)$ is correctly specified, what means

$$p(X,\alpha) = p(X) = \mathbb{E}[T|X] = \mathbb{E}[T|Y(1),X],$$

but the regression model is misspecified:

$$m_1(X,\beta_1) \neq \mathbb{E}(Y|T=1,X)$$

Then the second term in (43) is

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{T}{p(X)} - 1\right)(Y(1) - m_1(X,\beta_1))\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{T}{p(X)} - 1\right)(Y(1) - m_1(X,\beta_1))\middle| Y(1),X\right]\right] \\ &= \mathbb{E}\left[(Y(1) - m_1(X,\beta_1))\,\mathbb{E}\left[\left(\frac{T}{p(X)} - 1\right)\middle| Y(1),X\right]\right] \\ &= \mathbb{E}\left[(Y(1) - m_1(X,\beta_1))\left(\frac{\mathbb{E}[T|Y(1),X]}{p(X)} - 1\right)\right] \\ &= \mathbb{E}\left[(Y(1) - m_1(X,\beta_1))\left(\frac{p(X)}{p(X)} - 1\right)\right] = 0,\end{aligned}$$

and so $\hat{\mu}_{1,DR}$ estimates $E[Y(1)]$. Analogously $\hat{\mu}_{0,DR}$ estimates $E[Y(0)]$ and $\hat{\tau}_{DR}$ estimates the average treatment effect τ.

Situation 2: Model $m_1(X, \beta_1)$ is correctly specified and

$$m_1(X, \beta_1) = \mathbb{E}(Y|T = 1, X) = \mathbb{E}(Y(1)|X)$$

but the propensity score model is misspecified:

$$p(X, \alpha) \neq p(X) = \mathbb{E}[T|X].$$

In this case the second term in (43) is

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{T}{p(X,\alpha)} - 1\right)(Y(1) - \mathbb{E}(Y|T = 1, X))\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{T}{p(X,\alpha)} - 1\right)(Y(1) - \mathbb{E}(Y|T = 1, X))\middle|T, X\right]\right] \\
&= \mathbb{E}\left[\left(\frac{T}{p(X,\alpha)} - 1\right)\mathbb{E}\left[(Y(1) - \mathbb{E}(Y|T = 1, X))|T, X\right]\right] \\
&= \mathbb{E}\left[\left(\frac{T}{p(X,\alpha)} - 1\right)(\mathbb{E}\left(Y(1)|T, X\right) - \mathbb{E}(Y|T = 1, X))\right] = 0,
\end{aligned}$$

where the last equality is by unconfoundedness:

$$\mathbb{E}(Y|T = 1, X) = \mathbb{E}(Y(1)|T = 1, X) = \mathbb{E}(Y(1)|X) = \mathbb{E}(Y(1)|T, X).$$

Again $\hat{\mu}_{1,DR}$ estimates $E[Y(1)]$. Similarly $\hat{\mu}_{0,DR}$ estimates $E[Y(0)]$ and $\hat{\tau}_{DR}$ estimates τ.

If both models are correctly specified, clearly $\hat{\tau}_{DR}$ estimates τ. In case of the misspecification in both models, $\hat{\tau}_{DR}$ does not estimate τ and is therefore inconsistent (Davidian 2007).

B List of covariates

Table 4: Covariates

Variable	Description
personal characteristics	
age99	age at the program start or unemployed start in 1999
age3645	age 35-45
age46	age above 45
foreign	foreigner
famdiv	dummy divorced
famsin	dummy single
education	
edbasic	basic
edhbb	higher education
edaka	university
occupation	
jehc	health, education, culture
jjur	legal, administration
jtech	technology
jtou	tourism
jodl	other services
jtt	trade and transportation
jcon	construction
jporo	other production
last job industry	
indmf	manufacturing
indcon	construction
indtra	trade
indtou	tourism
indtc	transport, communication
indsd	business activities
indst	public administration
indehc	education, health
emp109	<10
emp499	100-499
emp1000	>499
last employment, wage	
lwage	wage in last job (log)
lwagemi	wage missing
lwagel	wage below median
federal states dummies	
wie	Vienna (regional dummy)
noe	Lower Austria
sbg	Salzburg
tir	Tyrol
bgl	Burgenland
stm	Styria
ktn	Carinthia
regional variables	

Continued on next page

Table 4 – continued

Variable	Description
bigr	big cities: Vienna, Graz, Klagenfurt, Linz, Salzburg, Innsbruck
tour	tourism region
indr	industrial area
fq99	share of foreign employment
albq99	unemployment rate
jalb99	share of youth unemployment
lalb99	rate of long-term unemployed (>6 months)
history 2 years before program start	
mb2	share employed
mal2	share unemployed
mdolf2	share out of labour force
ep_p2	number of programs
d_p2	duration of programs
k_dt	days on maternity leave
k5_dt	days on maternity leave 5 years before program start
btvf	days employed before program start
altvf	days unemployed before program start
dalb1	receipt of unemployment benefits
dblvm1	duration of last employment (in days)
blvm1	days from last job to program start
sstart	sample entry date
program type	
dm1lp	training program
dm2lp	active job search program
dm4lp	other program
entry in employment	
dmon4	dummy for beginning of unemployment in April
dmon5	dummy for beginning of unemployment in May
dmon6	dummy for beginning of unemployment in June
dmon7	dummy for beginning of unemployment in July
dmon8	dummy for beginning of unemployment in August

C Propensity score estimation

Table 5: Estimation of propensity score using probit model

	Female		Male	
	Active job-s.	**Training**	**Active job-s.**	**Training**
age99	-0.020439 ***	0.017332 **	-0.026191 ***	-0.01085
age3645	0.11943 ***	-0.015577	0.068304	0.010956
age46	0.12034	-0.18171 **	0.025458	-0.078069
foreign	-0.18856 ***	-0.28097 ***	-0.24977 ***	-0.43999 ***
fq99	-0.0034745	-0.066078 ***	-0.021762 **	-0.038208 ***
famdiv	0.21465 ***	0.1104	-0.036088	-0.034212
famsin	0.12668 ***	0.17775 ***	0.12213 ***	0.051369
jehc	-0.23663 **	-0.02938	-0.047663	0.1006
jjur	-0.068075	0.50224 ***	0.22857 **	0.47828 ***
jtech	0.037155	0.3939 **	0.18645 *	0.66493 ***
jtou	-0.42763 ***	-0.15833	-0.16325	0.0075297
jodl	-0.26723 **	-0.40674 ***	-0.0074707	-0.011812
jtt	-0.11573	0.055537	0.090419	0.13049
jcon	-0.41967	0.16674	-0.05692	-0.0078798
jporo	-0.077249	-0.018746	0.17062 *	0.10964
albq99	-0.0025027	-0.083357 ***	-0.054216 ***	-0.058839 ***
jalb99	0.019827 ***	0.01316	0.009346	0.020296 *
lalb99	-0.0046391	-0.0050289	-0.0037634	0.0088468 *
bigr	-0.089019 *	0.21679 ***	-0.14298 **	0.074243
tour	0.03947	-0.11249 *	-0.030525	-0.074026
indr	0.15609	0.019587	-0.47965 **	0.28758
wie	0.56676 ***	0.80558 ***	0.75465 ***	0.28631 *
noe	-0.017585	-0.25876 ***	-0.032587	-0.22081 ***
sbg	-0.25082 ***	-0.44261 ***	-0.26932 ***	-0.19842 **
tir	-0.11155 *	-0.80515 ***	-0.070576	-0.65724 ***
bgl	-0.29814 ***	-0.048583	-0.44065 ***	-0.073666
stm	0.51571 ***	-0.29036 ***	0.43036 ***	-0.25367 ***
ktn	-0.40699 ***	-0.13865 **	-0.34246 ***	-0.038512
edbasic	-0.032377	-0.08895 **	0.05126	-0.049079
edhbb	-0.27292 ***	-0.021746	-0.14053 **	-0.14038 **
edaka	-0.17471 **	-0.22047 ***	-0.10437	-0.23108 **
btvf	-0.00012197 **	3.6166e-005	-0.00018209 ***	-4.5599e-005
altvf	-9.0995e-005	0.00041011 ***	-0.00015715	0.00020728
mb2	-0.34729 *	0.34689 *	-0.00014281	-0.23147
mal2	1.0129 ***	0.35472 ***	0.99642 ***	0.50447 **
mdolf2	-0.65236 ***	0.035895	-0.83412 ***	-0.46649 *
dalb1	-0.49037	-0.5286	-0.22141	-0.40734
indmf	0.36013	0.27168	-0.29302	-0.19579
indcon	0.65974 *	0.50125	-0.29008	-0.19184
indtra	0.34714 *	0.42186 *	-0.29772	-0.40354
indtou	0.01647	0.67737 ***	-0.24733	-0.35449
indtc	0.20741	0.22046	-0.26347	-0.15502
indsd	0.21679	0.20002	-0.37819	-0.28071
indst	0.46436	0.36383	-0.7444 *	-0.87358 *

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Continued on next page

Table 5 – continued

	Female		Male	
	Active job-s.	**Training**	**Active job-s.**	**Training**
indehc	0.33234	0.44734	-0.63995 *	-0.61986
emp109	-0.054594 **	0.002631	0.014253	-0.01786
emp499	0.033415	-0.056332	0.093772 *	0.19401 ***
emp1000	-0.050855	0.29617 *	0.34135 **	0.06408
dblvm1	3.2902e-005 **	2.581e-005 *	3.5719e-005 **	1.9386e-005
lwage	0.048643	0.053018	0.04522	-0.098944
lwagemi	0.15417	0.50093	0.12999	-0.66299
lwageb	1.1918e-006	-1.3291e-005	-1.3931e-005	3.0117e-005 *
lwagediv	-0.0085529	0.015193	0.02158 *	0.0046592
wiebtvf	4.4157e-005 *	-9.409e-008	7.9711e-005 ***	-1.1067e-005
wiefamsin	0.041602	-0.10567 *	-0.052965	-0.011636
wiefamdiv	-0.04251	-0.13108 *	-0.097531	-0.022223
wiemb2	0.34199 ***	-0.044927	0.504 ***	0.061415
wieedbasic	0.054324	-0.0011896	-0.062678	-0.032108
wieedhbb	0.15918	0.10564	0.22794 **	0.18611 *
wieedaka	0.26453 **	0.17182	0.29492 **	0.0095352
blvm1	0.00024594 ***	0.00020543 ***	0.00036723 ***	5.031e-005
ep_p2	0.018414	0.15299 ***	0.046947	0.14235 ***
d_p2	6.1368e-005	-0.0003325	0.00029529	-0.00010206
dm2lp	0.069657	0.052262	0.68085 ***	0.19904
dmon4	-0.4653 ***	-0.33286 ***	-0.96574 ***	-0.52421 ***
dmon5	-0.15729 ***	-0.25814 ***	-0.29303 ***	-0.35228 ***
dmon6	-0.12478 ***	-0.25272 ***	-0.068223	-0.32922 ***
dmon7	-0.20726 ***	-0.4292 ***	-0.048303	-0.36756 ***
dmon8	-0.17662	-0.57706 ***	-0.1474	-0.25515 *
wieblvm1	-0.00016267 ***	-0.0001418 ***	-0.00024994 ***	5.3723e-005
stmblvm1	-0.00016631 ***	-0.0001725 ***	-0.0003056 ***	4.7159e-005
mb2blvm1	0.0048887 ***	0.0035328 ***	0.0055088 ***	0.003684 ***
edbasicblvm1	-2.1069e-006	-6.3212e-005 *	-0.00016179 ***	-3.3762e-005
mdolf2blvm1	-0.00021434 ***	3.1387e-005	-0.00017105	-9.1492e-005
dm2lpblvm1	-7.806e-005	0.00021863	-0.00050711	0.00024974
mb2age99	-0.0024296	-0.0021125	-0.013383 **	5.9294e-005
btvfage99	3.2726e-006 **	-5.8181e-007	5.1556e-006 ***	1.8085e-006
altvfage99	1.0329e-006	-1.1737e-005 ***	3.9011e-006	-7.8066e-006 **
dalb1lwage	0.062011	0.076555	0.027314	0.053359
dalb1lwagemi	0.29323	0.43774	0.038062	0.32089
mb2indst	-0.48275 ***	-0.32027 **	-0.4305 *	-0.19762
mb2indehc	0.12999	-0.14067	0.25258	0.38571
mdolf2bigr	0.087354	0.022704	0.15347	0.080896
mdolf2lwagemi	0.67272 ***	0.12887	0.90444 ***	0.33807 **
mdolf2indtc	0.28029	0.25886	0.47602 **	0.092132
mdolf2tour	-0.0035803	0.17216	0.40679 *	0.20521
age99indmf	-0.0033369	-0.0054374	0.014858 **	0.0087726
age99indtra	-0.0040041	-0.0080937	0.017073 ***	0.015787 **
age99indtou	0.0014072	-0.018404 **	0.011471	0.012449
age99indtc	-0.0025791	-0.007808	0.0099227	0.0030106

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Continued on next page

Table 5 – continued

	Female		Male	
	Active job-s.	**Training**	**Active job-s.**	**Training**
age99indsd	-0.0016432	-0.002934	0.015281 **	0.0097267
age99indst	-0.0050633	-0.0054222	0.026385 **	0.025271 **
age99indehc	-0.0074748	-0.0084604	0.016336	0.012567
age99indcon	-0.016536	-0.010959	0.0060087	0.0050848
dmon4indmf	0.17614	-0.070287	0.49285 ***	0.088902
dmon4indtra	0.20682 *	0.037096	0.31375 *	0.21724
dmon4indsd	-0.22677	0.15894	0.42155 **	0.12091
dmon4indehc	0.06757	-0.078919	0.59592 *	
dmon4indcon	0.40718	0.15634	0.46621 **	0.033956
dmon8indmf	-0.22046	0.03339	-0.31447 *	-0.073148
dmon8indtra	-0.37191 **	0.19707	-0.53963 ***	-0.12734
dmon8indtou	-0.22425	0.042746	-0.37365 **	-0.39217 *
dmon8indtc	-0.70195 ***	0.22887	-0.52276 ***	-0.083298
dmon8indsd	-0.42488 ***	0.10828	-0.443 ***	-0.025092
dmon8indst	-0.5453 **	0.091265	-0.50633	-0.23373
dmon8indehc	-0.26307	0.15746	-0.38001	-0.048732
dmon8indcon	-0.24868	0.054023	-0.26344	-0.033882
dmon5indtou	-0.29309 ***	-0.42817 ***	-0.24274 **	-0.5904 ***
dmon5indst	-0.071877	0.070431	0.59457 ***	-0.30253
dmon6indtra	0.035237	-0.15521 *	-0.17045 **	0.018079
dmon6indtou	-0.12201	-0.40151 ***	-0.26448 **	-0.21792
dmon7indehc	0.14062	-0.19319	0.40454 **	0.6265 ***
dmon7indcon	0.24219	0.20514	0.19401 **	-0.13824
dmon4indr	0.099055	0.12896	0.12999	0.114
dmon5indr	-0.096841	0.089914	0.25031 ***	-0.088967
dmon6indr	-0.25249 ***	0.044469	-0.056945	0.088712
dmon7indr	-0.33897 ***	-0.15061 *	-0.28857 ***	-0.085882
dmon6dm1lp	0.17682	-0.26982	-0.21701	0.15497
dmon7dm1lp	-0.1502	-0.31405	-0.27378	-0.051653
dmon5dm2lp	0.5114	-0.83441	-0.46681	-0.05478
lwageljtou	0.06483	-0.0038595	0.03934	-0.073788
lwagelmb2	0.095668	-0.18466 *	-0.37916 ***	0.016467
dalb1dm2lp	-0.47085	0.25869	-0.57907 *	0.083958
wiedalb1	0.31702 ***	-0.07394	0.32042 ***	0.10847 *
indralbq99	-0.011151	0.0056817	0.048202 *	-0.0066864
indrfq99	0.0078006	-0.011302	0.026305 **	-0.023241 **
indredbasic	0.098006 *	0.060452	-0.1511 **	-0.14112 **
d_p2dm1lp	0.0004956	0.00047572	0.0012174 **	-0.0001426
lwagel	-0.040647	0.15538 *	0.22815 ***	-0.090561
k_dt	-9.4864e-006	-4.905e-005		
k5_dt	1.1877e-006	0.00015629 **		
dm1lp	-0.01002	0.079692	0.0032667	0.36421 ***
dm4lp	0.1299 ***	0.17926 ***	0.012189	0.22199 ***
sstart	2.9902e-005	2.2643e-005	5.0624e-005 **	3.3656e-005
_cons	-1.7696 ***	-2.0607 ***	-1.5268 **	-0.70184
N	30430	28453	25958	23343

D Matching estimation results

	$\hat{\tau}_{\mathbf{MATCH}}$
Female	
Active job search programs	26.739
	(19.699)
Training programs	20.035
	(20.622)
Male	
Active job search programs	-27.338
	(23.985)
Training programs	-64.187**
	(25.634)

Table 6: Average treatment effects estimated using nearest neighbor matching. In parentheses heteroskedasticity-consistent analytical standard errors according to Abadie and Imbens (2006). *, ** and *** denote significance at 10%, 5% and 1%.

E Matlab code

```
%%===================== FEMALE ajsp program =====================

global T X y
load('f_ajsp.mat')
y = obj4c;
n = size(T,1);
X= [ ones(n,1) age99 age3645 age46 foreign fq99 famdiv famsin ...
jehc jjur jtech jtou jodl jtt jcon jporo albq99 jalb99 lalb99 ...
bigr tour indr wie noe sbg tir bgl stm ktn  edbasic edhbb edaka ...
btvf altvf mb2 mal2  mdolf2 dalb1 indmf indcon indtra indtou ...
indtc indsd indst indehc emp109 emp499 emp1000 dblvm1 lwage ...
lwagemi lwageb lwagediv wiebtvf wiefamsin wiefamdiv wiemb2 ...
wieedbasic wieedhbb wieedaka  blvm1  ep_p2 d_p2  dm2lp  dmon4 ...
dmon5 dmon6 dmon7 dmon8   wieblvm1  stmblvm1 mb2blvm1 ...
edbasicblvm1 mdolf2blvm1 dm2lpblvm1 mb2age99  btvfage99 ...
altvfage99 dalb1lwage dalb1lwagemi mb2indst mb2indehc mdolf2bigr ...
mdolf2lwagemi mdolf2indtc  mdolf2tour age99indmf age99indtra ...
age99indtou age99indtc age99indsd age99indst age99indehc ...
age99indcon dmon4indmf dmon4indtra  dmon4indsd  dmon4indehc ...
dmon4indcon dmon8indmf dmon8indtra dmon8indtou dmon8indtc ...
dmon8indsd dmon8indst dmon8indehc  dmon8indcon    dmon5indtou ...
dmon5indst dmon6indtra dmon6indtou dmon7indehc  dmon7indcon ...
dmon4indr dmon5indr dmon6indr dmon7indr dmon6dm1lp dmon7dm1lp ...
dmon5dm2lp lwageljtou lwagelmb2 dalb1dm2lp wiedalb1 indralbq99 ...
indrfq99 indredbasic d_p2dm1lp lwagel k_dt k5_dt dm1lp dm4lp ...
sstart];
k = size(X,2);
init = [initalpha initmu1 initmu0];

%% -------------------------- Trimming --------------------------
% estimation of propensity score for trimming
thet = fsolve(@psi_ps1,init');
alpha = thet(1:(end-2));

prop=normcdf( X*alpha' );
prop1=prop(T==1);
prop0=prop(T==0);

X((prop<min(prop1))|(prop>max(prop0)),:)=[];
y((prop<min(prop1))|(prop>max(prop0)))=[];
T((prop<min(prop1))|(prop>max(prop0)),:)=[];
n = size(T,1);

%% ---------------------------- IPW1 ----------------------------
[thet_ps1,fval,exitflag,output,Aps1] = fsolve(@psi_ps1,init');

alpha = thet_ps1(1:(end-2));
mu1 = thet_ps1(end-1);
mu0 = thet_ps1(end);
% tau_ps1  (tau_IPW1)
tau_ps1 = mu1-mu0

% matrix V_ps1  (B_IPW1)
v11= 1/(n)  * X' *  ( X.*(((T - normcdf( X*alpha) ) ./   ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha)).^2*ones(1,k)) );
psi2 = T.*y ./ normcdf( X*alpha )- mu1;
```

```
psi3 = (1-T).*y ./ ( 1-normcdf( X*alpha ) ) - mu0;

v12= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi2));
v13= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi3));
v33= 1/(n) *[ psi2'*psi2, psi2'*psi3; psi2'*psi3, psi3'*psi3  ];
V_ps1 = [v11,v12,v13; [v12';v13'],v33];

% estimation of SE for theta_ps1
thetaavar = inv(Aps1)*V_ps1*inv(Aps1)';
muavar = thetaavar((end-1):end,(end-1):end);
tauavar = muavar(1,1)+muavar(2,2)-2*muavar(1,2);
% asymptotic SE for tau_ps1 (tau_IPW1)
tausd_ps1 = sqrt(tauavar/n)
t_ps1 = tau_ps1/tausd_ps1

%% ---------------------------- IPW2 ----------------------------
[thet_ps2,fval,exitflag,output,Aps2] = fsolve(@psi_ps2,init');

alpha = thet_ps2(1:(end-2));
mu1 = thet_ps2(end-1);
mu0 = thet_ps2(end);
% tau_ps2 (tau_IPW2)
tau_ps2 = mu1-mu0

% matrix V_ps2 (B_IPW2)
v11= 1/(n) * X' * ( X.*(((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha)).^2*ones(1,k)) );
psi2 = T.*(y- mu1) ./ normcdf( X*alpha );
psi3 = (1-T).*(y- mu0)./ ( 1-normcdf( X*alpha ) );
v12= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi2));
v13= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi3));
v33= 1/(n) *[ psi2'*psi2, psi2'*psi3; psi2'*psi3, psi3'*psi3  ];
V_ps2 = [v11,v12,v13; [v12';v13'],v33];

% estimation of SE for theta_ps2
thetaavar = inv(Aps2)*V_ps2*inv(Aps2)';
muavar = thetaavar((end-1):end,(end-1):end);
tauavar = muavar(1,1)+muavar(2,2)-2*muavar(1,2);
% asymptotic SE for tau_ps2 (tau_IPW2)
tausd_ps2 = sqrt(tauavar/n)
t_ps2 = tau_ps2/tausd_ps2

%% ---------------------------- IPW3 ----------------------------
[thet_ps3,fval,exitflag,output,Aps3] = fsolve(@psi_ps3,init');

alpha = thet_ps3(1:(end-2));
mu1 = thet_ps3(end-1);
mu0 = thet_ps3(end);
% tau_ps3 (tau_IPW3)
tau_ps3 = mu1-mu0
```

```
% matrix V_ps3 (B_IPW3)
v11= 1/(n) * X' * ( X.*(((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha)).^2*ones(1,k)) );

    eta1_nom = -1/n * ( T.*(y-mu1))' * ...
            (ones( n,1)./(normcdf( X*alpha)).^2);
    eta1_den = 1/n * ((T-normcdf( X*alpha)).^2 )' * ...
            (ones( n,1)./(normcdf( X*alpha)).^2);
    eta1 = eta1_nom/eta1_den;

    eta0_nom = -1/n * ( (1-T).*(y-mu0) )' * ...
            (ones( n,1)./(1-normcdf( X*alpha)).^2);
    eta0_den = 1/n * ( (T-normcdf( X*alpha)).^2 )' * ...
            (ones( n,1)./(1-normcdf( X*alpha)).^2);
    eta0 = eta0_nom/eta0_den;

psi2 = T.*(y- mu1) ./ normcdf( X*alpha ) ...
    + eta1 * (T- normcdf( X*alpha )) ./ normcdf( X*alpha );
psi3 = (1-T).*(y- mu0) ./ (1-normcdf( X*alpha )) ...
    - eta0 * (T- normcdf( X*alpha )) ./ (1-normcdf( X*alpha ));

v12= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha).*psi2));
v13= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha).*psi3));
v33= 1/(n) *[ psi2'*psi2, psi2'*psi3; psi2'*psi3, psi3'*psi3  ];
V_ps3 = [v11,v12,v13; [v12';v13'],v33];

% estimation of SE for theta_ps3
thetaavar = inv(Aps3)*V_ps3*inv(Aps3)';
muavar = thetaavar((end-1):end,(end-1):end);
tauavar = muavar(1,1)+muavar(2,2)-2*muavar(1,2);
% asymptotic SE for tau_ps3 (tau_IPW3)
tausd_ps3 = sqrt(tauavar/n)
t_ps3 = tau_ps3/tausd_ps3

%% ----------------------------- DR -----------------------------
thetainit = [init(1:k) initb1 initb0 init(end-1:end)];

[thet_dr,fval,exitflag,output,Adr] = fsolve(@psi_dr,thetainit');

alpha = thet_dr(1:k);
beta1 = thet_dr(k+1:k+k);
beta0 = thet_dr(2*k+1:3*k);
mu1 = thet_dr(end-1);
mu0 = thet_dr(end);
tau_dr = mu1-mu0

% matrix V_dr (B_DR)
v11= 1/(n) * X' * ( X.*(((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha)).^2*ones(1,k)) );
v12= -1/(n) * X' * ( X.*(((T - normcdf( X*alpha) ) ./  ...
           (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) .* ...
           normpdf(X*alpha)).*( T .* (2*(y-X*beta1)) )*ones(1,k)));
```

```
v13= -1/(n) * X' * ( X.*(((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) .* ...
            normpdf(X*alpha)).*( (1-T).*(2*(y-X*beta0)))*ones(1,k)));

psi4 = T.*y ./ normcdf( X*alpha ) - (T./ normcdf( X*alpha )-1).* ...
       (X*beta1) - mu1;
psi5 = (1-T).*y ./ (1-normcdf( X*alpha )) - ((1-T)./ ...
      (1-normcdf( X*alpha )) -1).* (X*beta0) - mu0;
v14= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi4));
v15= 1/(n) * X' * (((T - normcdf( X*alpha) ) ./  ...
             (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
             .* normpdf(X*alpha).*psi5));
v22= 1/(n) * X' * ( X.*(( T .* (2*(y-X*beta1) )).^2*ones(1,k)) );
v23= 1/(n) * X' * ( X.*(( T .* (2*(y-X*beta1) )).* ...
     ( (1-T) .* (2*(y-X*beta0)) )*ones(1,k)) );
v24= -1/(n) * X' * ((( T .* (2*(y-X*beta1) )).*psi4));
v25= -1/(n) * X' * ((( T .* (2*(y-X*beta1) )).*psi5));
v33= 1/(n) * X' * ( X.*(( (1-T) .* (2*(y-X*beta0))).^2*ones(1,k)) );
v34= -1/(n) * X' * ((( (1-T) .* (2*(y-X*beta0)) ).*psi4));
v35= -1/(n) * X' * ((( (1-T) .* (2*(y-X*beta0)) ).*psi5));
v55= 1/(n) *[ psi4'*psi4, psi4'*psi5; psi4'*psi5, psi5'*psi5  ];
V_dr = [v11, v12, v13,v14,v15; ...
        v12',v22, v23,v24,v25; ...
        v13',v23',v33,v34,v35; ...
        [v14';v15'],[v24';v25'],[v34';v35'],v55];

% estimation of SE for theta_dr
thetaavar = inv(Adr)*V_dr*inv(Adr)';
muavar = thetaavar((end-1):end,(end-1):end);
tauavar = muavar(1,1)+muavar(2,2)-2*muavar(1,2);
% asymptotic SE for tau_ps3 (tau_IPW3)
tausd_dr = sqrt(tauavar/n)
t_dr = tau_dr/tausd_dr

%% --------------------- Function for IPW1 ----------------------
moment conditions for IPW1
function mpsi = psi_ps1(theta)

    global T X y
    k = size(X,2);
    n = size(T,1);
    alpha = theta(1:(end-2));
    mu1 = theta(end-1);
    mu0 = theta(end);
    mpsi1 = (1/n)* X' * ((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha));
    psi2 = T.*y ./ normcdf( X*alpha )- mu1;
    psi3 = (1-T).*y ./ ( 1-normcdf( X*alpha ) ) - mu0;
    mpsi2 = mean(psi2);
    mpsi3 = mean(psi3);

    mpsi = [mpsi1' mpsi2 mpsi3];

end
```

```
%% ---------------------- Function for IPW2 -----------------------
% moment conditions for IPW2
function mpsi = psi_ps2(theta)

    global T X y
    k = size(X,2);
    n = size(T,1);
    alpha = theta(1:(end-2));
    mu1 = theta(end-1);
    mu0 = theta(end);

    mpsi1 = (1/n)* X' * ((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha));
    psi2 = T.*(y- mu1) ./ normcdf( X*alpha );
    psi3 = (1-T).*(y - mu0) ./ ( 1-normcdf( X*alpha ) );
    mpsi2 = mean(psi2);
    mpsi3 = mean(psi3);

    mpsi = [mpsi1' mpsi2 mpsi3];
end

%% ---------------------- Function for IPW3 -----------------------
% moment conditions for IPW3
function mpsi = psi_ps3(theta)

    global T X y
    k = size(X,2);
    n = size(T,1);
    alpha = theta(1:(end-2));
    mu1 = theta(end-1);
    mu0 = theta(end);

    mpsi1 = (1/n)* X' * ((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha));
    eta1_nom = -1/n * ( T.*(y-mu1) )' * ...
            (ones( n,1)./(normcdf( X*alpha)).^2);
    eta1_den = 1/n * ((T-normcdf( X*alpha)).^2 )' * ...
            (ones( n,1)./(normcdf( X*alpha)).^2);
    eta1 = eta1_nom/eta1_den;

    eta0_nom = -1/n * ( (1-T).*(y-mu0))' * ...
            (ones( n,1)./(1-normcdf( X*alpha)).^2);
    eta0_den = 1/n * ( (T-normcdf( X*alpha)).^2 )' * ...
            (ones( n,1)./(1-normcdf( X*alpha)).^2);
    eta0 = eta0_nom/eta0_den;

    mpsi2 = 1/n*(T.*(y- mu1))'*(ones( n,1)./(normcdf(X*alpha))) ...
           + 1/n * eta1 * (T - normcdf( X*alpha))' * ...
           (ones( n,1)./(normcdf( X*alpha)));
    mpsi3 = 1/n * ((1-T).*(y- mu0))' * (ones( n,1)./ ...
           (1-normcdf( X*alpha))) - ...
           1/n * eta0 * (T - normcdf( X*alpha))' * ...
           (ones( n,1)./(1-normcdf( X*alpha)));

    mpsi = [mpsi1' mpsi2 mpsi3];
end
```

```
%% ---------------------- Function for DR -----------------------
% moment conditions for DR
function mpsi = psi_dr(theta)

    global T X y
    k = size(X,2);
    n = size(T,1);
    alpha = theta(1:k);
    beta1 = theta(k+1:k+k);
    beta0 = theta(2*k+1:3*k);
    mu1 = theta(end-1);
    mu0 = theta(end);

    mpsi1 = (1/n)* X' * ((T - normcdf( X*alpha) ) ./  ...
            (normcdf( X*alpha ).*( 1-normcdf( X*alpha ))) ...
            .* normpdf(X*alpha));
    mpsi2 = -(1/n)*X' * ( T .* (2*(y-X*beta1)) );
    mpsi3 = -(1/n)*X' * ( (1-T) .* (2*(y-X*beta0)) );
    psi4 = T.*y ./ normcdf( X*alpha ) - ...
          (T./ normcdf( X*alpha )-1).* (X*beta1) - mu1;
    psi5 = (1-T).*y ./ (1-normcdf( X*alpha )) - ...
           ((1-T)./ (1-normcdf( X*alpha ))-1).* (X*beta0) - mu0;
    mpsi4 = mean(psi4);
    mpsi5 = mean(psi5);

    mpsi = [mpsi1' mpsi2' mpsi3' mpsi4 mpsi5];
end
```