

VATT-KESKUSTELUALOITTEITA
VATT DISCUSSION PAPERS

438

VARYING BIASES
IN THE MATCHING
ESTIMATES:
EVIDENCE FROM
TWO RANDOMIZED
JOB SEARCH
TRAINING
EXPERIMENTS

Kari Hämäläinen
Roope Uusitalo
Jukka Vuori

ISBN 978-951-561-757-6 (nid.)

ISBN 978-951-561-758-3 (PDF)

ISSN 0788-5016 (nid.)

ISSN 1795-3359 (PDF)

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Arkadiankatu 7, 00100 Helsinki, Finland

Email: etunimi.sukunimi@vatt.fi

Oy Nord Print Ab

Helsinki, January 2008

HÄMÄLÄINEN, KARI – UUSITALO, ROOPE – VUORI, JUKKA: VARYING BIASES IN THE MATCHING ESTIMATES: EVIDENCE FROM TWO RANDOMIZED JOB SEARCH TRAINING EXPERIMENTS. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 2008, (C, ISSN 0788-5016 (nid.), ISSN 1795-3359 (PDF No 438). ISBN 978-951-561-757-6 (nid.), ISBN 978-951-561-758-3 (PDF).

Abstract: We compare various matching estimators to the results of two randomized field experiments that evaluate employment effects of job search training programs. We find that commonly used non-experimental matching estimators tend to over-estimate the program effects and that the bias differs between programs owing to different procedures used in selecting the program participants. The bias is larger when participation is voluntary than when caseworkers assign the applicants to training.

Keywords: Job search training, field experiments, matching, unemployment

JEL codes: C93

HÄMÄLÄINEN, KARI – UUSITALO, ROOPE – VUORI, JUKKA: VARYING BIASES IN THE MATCHING ESTIMATES: EVIDENCE FROM TWO RANDOMIZED JOB SEARCH TRAINING EXPERIMENTS. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 2008, (C, ISSN 0788-5016 (nid.), ISSN 1795-3359 (PDF No 438). ISBN 978-951-561-757-6 (nid.), ISBN 978-951-561-758-3 (PDF).

Tiivistelmä: Vertaamme kaltaistamistuloksia kahteen satunnaistettuun kenttäkokeeseen, joissa on arvioitu työnhakukoulutuksen työllisyysvaikutuksia. Tulosten perusteella ei-kokeelliset arviointitutkimukset yliarvioivat toimenpiteen todellisia vaikutuksia. Harhan suuruus riippuu osallistujien valikoitumisesta. Osallistumiseltaan vapaaehtoisessa koulutuksessa harha on suurempi kuin siinä tapauksessa, että työvoimaneuvojat määräävät henkilöt työnhakukoulutukseen.

Asiasanat: Työn hakukoulutus, kenttäkoe, kaltaistaminen, työttömyys

Contents

1. Introduction	1
2. Job search training experiments	4
3. Data collection	7
4. Results	10
4.1 Long-term effects on job search training	11
4.2 Comparison of experimental and non-experimental estimates	13
5. Conclusion	17
References	18
Appendix	20

1. Introduction¹

Job search training is a common element of labor market programs in many European countries. These training courses are aimed at improving job search skills and provide help in writing a job application and cv, contacting employers and preparing for a job interview. The courses also attempt to help unemployed to recognize their skills and act as support groups that prepare for setbacks in the job search process. The policy is attractive because the training courses are rather short, typically lasting between one and two weeks, and job search training is, therefore, potentially cost-effective way of getting the unemployed back to work.

In 1996, the Finnish Institute of Occupational Health carried out a demonstration project “Työhön” that evaluated the employment effects of a job search training program using a randomized field experiment. The participants were selected by randomly allocating a half of the applicants to a training course and a half to a control group that did not participate in training. Three years after the first experiment job search training courses were added as a standard element to more extensive labor market training programs. In connection to this labor market policy reform, another randomized field experiment was carried out in nineteen employment offices across the country. Again participants were randomly chosen, this time by assigning two thirds of the eligible applicants to a job search training course and leaving one third in the control group.

Both courses were based on a group intervention lasting for one week. In both cases over one thousand unemployed took part in the experiment. The main difference between the programs was that the demonstration project followed closely the intervention design with the researchers monitoring its implementation, while the latter courses were usual job search training courses that applied various training methods and that were organized by the local employment offices. Another key difference had to do with the selection process. Participation in the demonstration project was entirely voluntary and the participants were recruited by advertising the courses at local employment offices and in local newspapers. In the second experiment the program participants were first selected by the employment offices using their usual procedures and after being already assigned to a program asked if they were willing to take part in a randomized experiment.

The employment effects of these job search training experiments have been documented previously by Vuori, Silvonen, Vinokur & Price (2002) and Malmberg-Heimonen and Vuori (2005). These studies used data from the follow-up surveys conducted six months after program participation. Vuori and Silvonen

¹ We would like to thank the Finnish Ministry of Labor for financial support and Tiina Salokangas, Elina Nykyri and Miikka Rokkanen for their excellent research assistance.

(2005) also report results based on a second follow-up study two years after the experiment. None of these studies found significant main effects, but Vuori and Silvonen (2005) report positive employment effects among participants who are in risk of depression.

Most previous empirical evidence on the effects of job search training comes from programs that offer a package of treatments. For example, Hotz et al. (2006) re-evaluate experimental results from the California GAIN program. Dolton and O'Neill (2002) evaluate the Restart program and Blundell et al. (2004) the New Deal program in the UK. Both programs include job search training but also other elements such as monitoring, basic skills courses or subsidized job placement. Van den Berg and van der Klaauw (2006) present results from a randomized social experiment in the Netherlands but also this program included not only job search assistance but also tighter monitoring. With the exemption of the Dutch study, the results show that job search assistance, possibly combined with other elements, improves employment prospects of the participants.

In this paper we combine data from the two Finnish experiments to administrative register data on the participants. This allows us to trace the effects of job search training on a monthly basis from the date of randomization up to six years after participation. We also collect data from a random sample of persons that were unemployed during the experiments and use these data to calculate various non-experimental matching estimators of the program effects.

Our paper contributes to the literature that compares non-experimental estimates to the experimental results, a tradition started by the LaLonde (1986) analysis of National Supported Work Program. Much of the more recent research has focused on another large scale randomized experiment evaluating the effects of Job Training Partnership Act described in detail by Bloom et al. (1997). A thorough analysis that compares the bias of several non-experimental estimators of the program effect is presented by Heckman, Ichimura, Smith and Todd (1998). Their conclusion is that bias in the non-experimental estimates can be substantially reduced by using data on non-participants from the same labor markets, administering the same survey instrument to both groups and by including information on recent labor market histories. Still although the non-experimental methods reduce the bias they do not eliminate it highlighting the benefits of randomized trials.

However, so far the most experimental evidence on the effects of labor market programs comes from the U.S. The ability of non-experimental methods to replicate experimental results also depends on the selection process into the programs. These processes vary across countries and across different programs within individual countries. Michalopoulos, Bloom and Hill (2004) and Agodini and Dynarski (2004) also argue that non-experimental method's ability to replicate experimental results may depend on whether the program under

evaluation is mandatory or not. By comparing the results based on two different experiments that differ in how participants to the program are selected, we will be able to evaluate how different selection processes influence the bias in the non-experimental procedures.

The rest of this paper is organized as follows. Section 2 describes the contents of the intervention and the design of randomized field experiments. In Section 3 we report how the participants of the experiment were traced from the register data and compare the self-reported outcomes to the outcomes deduced from the administrative registers. In Section 4 we report the long-term outcomes of the job search training programs and compare these results to various non-experimental matching estimators. Section 5 concludes with some final comments.

2. Job search training experiments

The Työhön experiment 1996-1997

The Työhön Job Search Program was based on the JOBS method designed at the Michigan Prevention Research Center as a preventive intervention for recently unemployed job seekers to facilitate their return to labor market and to prevent the negative mental health consequences of unemployment (Vinokur, Price and Schul, 1995).

In contrast to the U.S. version, the Finnish program also recruited long-term unemployed to job search training and did not screen for risk of depression. The Työhön-program aimed at improving job search skills and provided help in writing a job application and cv, contacting employers and preparing for a job interview. It consisted of five half-day sessions organized in groups of 6 to 17 unemployed participants. Each group had two trainers, who were themselves recruited from unemployed job seekers and trained for two months.

Recruitment and randomization

The trainers recruited the participants from the area of four employment offices in southwestern Finland using various methods including newspaper ads, presentations at information sessions for recently unemployed and direct contacts by mail. The employment offices were not directly involved in recruitment which also implied that no one could be assigned to the program but that participation was entirely voluntary. The recruiters estimated that they personally contacted 5000 unemployed. Eventually, 1471 agreed to participate in the experiment. Almost 90 percent of the participants were unemployed at the time when they were recruited. The median duration of elapsed unemployment was five months. Others had received a termination notice or were searching for a new job for some other reason. Only unemployed were eventually allowed to participate in the experiment.

The 1261 eligible unemployed who completed the pre-test questionnaire were randomized into treatment ($n = 629$) and control groups ($n = 632$). Those randomized into the treatment group were invited to participate in the program that typically started about three weeks after randomization. Of these 70 percent started in the program. The most common reason of dropping out was finding a job before the program started.

Job search training experiment 1999

The second experiment differed from the earlier demonstration program in a number of ways. These job search training programs were part of the usual re-employment services of the employment offices. The participants took part in one of the 137 job search training groups organized by the participating 19 local employment offices. In most cases the groups included also other participants than those taking part in the randomized experiment. Including these other participants the group size was between 8 and 20. Program length varied from 3 to 10 days (median 5) and hours per day from 3 to 7 (median 5). The trainers were career advisors employed at the employment offices. There was also variation in the content and the quality of training. Some of the courses relied on the Työhön -program while others used other methods or the trainer could not specify the method used. Variable quality was also reflected in participant evaluations conducted immediately after the training course. Variation in the quality assessments across courses was much larger than in the strictly structured Työhön -program implemented three years earlier.

Recruitment and randomization

The participants to job search training were selected by the caseworkers at employment offices. Only those who had no previous job search training but who, according to the caseworker assessment, could benefit from job search training were assigned to the program. Since the job search training groups were part of the usual services of the employment offices, refusing to participate in training could lead to sanctions and reductions of unemployment benefits though we have no data to verify whether sanctions were imposed.

After being screened by the caseworkers, the selected unemployed were approached by mail, or personally in connection to their regular visits at employment offices, and asked if they were willing to participate in an experiment evaluating the effects of job search training. They were told that program participants would be randomly assigned to treatment and control groups. Those who wanted to attend job search training immediately but did not want to participate in the experiment were placed to training programs and were not part of the experiment. Only those 1017 unemployed who agreed to randomization procedure and returned the baseline questionnaire were then randomly allocated to treatment (677) and control (338) groups. The treatment group was invited to a job search training program starting immediately. Those assigned to control group were excluded from job search training programs for six months but might participate later.

Since the 1996-7 program was a pilot project the only way to end up in the training course was to participate in the experiment. In the 1999 program, those

who did not participate in the experiment might still attend the job search training or other employment services. We will be estimating the effects of intention to treat counting these participants as members of the comparison group.

3. Data collection

In order to find those who took part in the experiment from administrative registers we first had to attach a personal id-number to each person. Using information on the date of birth together with names and addresses stored by Finnish Institute of Occupational Health we were able to find all but four participants from the Population Register. Also for almost everyone, we could be certain that we found a right person. Only in five cases in the first experiment and three cases in the second experiment we could not find exact matches of current or previous addresses.

These data that now only included a person-id and limited amount of basic information (treatment status, date of participation, employment status in the follow-up survey) was then sent to Statistics Finland. There the experimental data was linked to Employment Statistics, an administrative register database containing information on all persons living in Finland gathered from more than twenty different registers. For our purposes the most important information sources were the Job Seekers register of the Ministry of Labor containing all unemployment spells, the Tax Register containing annual earnings and the Pension Register reporting all employment spells that count for future pensions.

Statistics Finland attached information on the unemployment and employment spells, annual earnings and various background variables to each participant's records. We used data on employment and unemployment spells and the information on the labor market status in the end of each year to create an indicator of labor market status on a monthly basis. In most of the analysis we classify the individuals into five categories: employed, unemployed, in subsidized job, in labor market training and outside of the labor force. Data was collected from 1993 to 2003 allowing us to track the monthly labor market status both before experiment and up to six years after the first, and four years after the second experiment. We also collected a random sample of 5000 persons who were unemployed during at least one of the experiments from the same register to be used as a non-experimental comparison group. Some of these comparison group members probably participated job search training courses but we have no data to verify this. Finally, to comply with the data protection requirements, Statistics Finland removed the person-id and returned the data to us.

In the first experiment 78 percent of the participants were women. The average age was 37 and average duration of current unemployment spell four months. 78 percent had completed at least secondary education and 14 percent had a tertiary degree. Vuori et. al. (2002) compare the participants to average unemployed and note that both in terms of education vocational background the program participants were clearly in a stronger position in the labor market than the average unemployed in the same region.

The participants in the second experiment were less often women (60%) and had been unemployed for a longer period (average duration of current spell was 7.2 months) than in the participants of the first experiment. Also the education level was slightly lower, 74 percent had at least secondary education. Descriptive statistics of the experimental control group and the matched comparison groups are reported in detail in the Appendix A.

Differences between register and survey data

To verify data quality, we first compared register data to the survey responses. In the follow-up surveys conducted six months after the training program, both the treatment and comparison group members were asked about their current labor market status. We re-classified these answers into four categories: 1) employed, 2) unemployed, 3) in labor market training and 4) outside the labor force. The first category includes also those in subsidized jobs, so we combined open employment and subsidized jobs also in the register data. The second and third categories are easily defined both from register and from the survey data. The last category included all others, according to the survey responses these were most often mothers of small children or full-time students.

The results based on survey responses may well be different from the results based on register data. The program participants have typically a rather unstable job market attachment and the survey responses may vary according to the exact reference date. In some cases it is unclear how a labor market status should be defined, for example, if a person is employed but works only a few hours per week, and does not necessarily work at all during the reference week. To assess these differences we cross-tabulated the labor market status six months after the Työhön-program according to the survey and the register data in the upper part of Table 1. The lower part of Table 1 reports the same for the 1999 experiment.

According to the results in Table 1, the register data and the survey data are in most cases consistent, especially so for those who are unemployed or in labor market training. We also usually classify employment in the same way based on the register data and the survey responses. Results are less consistent for those who are outside the labor force. For this group the quality of register data is questionable. For example, maternity leave spells are not included in the data and many mothers who are on leave will be coded as employed. Also full-time students may be coded as employed even if they work just a few hours per week. One might also note that non-response rate is very small in the first follow-up survey (2.8%). In the latter experiment non-response is much higher (13.8%) but there is no clear sign that non-response would be strongly correlated with the labor market status.

Table 1. *Labor market status 6 months after treatment*

Työhön-program 1996/7						
Labor market status according to the register data						
	Employed	Unem- ployed	In labor market training	Outside of the LF	Missing	Total
Labor market status in the survey data						
Employed	455 (76.7 %)	34 (7.6 %)	2 (1.4 %)	9 (14.3 %)	1 (25.0 %)	501 (40.0%)
Unemployed	71 (12.0 %)	374 (83.9 %)	9 (6.1 %)	20 (31.8 %)	1 (25.0 %)	475 (37.9%)
In labor market training	26 (4.4 %)	21 (4.7 %)	134 (90.5 %)	9 (14.3 %)	0 (0.0 %)	190 (15.2%)
Outside of the labor force	23 (3.9 %)	9 (2.0 %)	0 (0.0 %)	21 (33.3 %)	0 (0.0 %)	53 (4.2%)
Missing	18 (3.0 %)	8 (1.8 %)	3 (2.0 %)	4 (6.4 %)	2 (50.0)	35 (2.8)
Total	593 (100.0 %)	446 (100.0 %)	148 (100.0 %)	63 (100.0 %)	4 (100.0 %)	1254 (100.0%)

Job search training 1999						
Labor market status according to the register data						
	Employed	Unem- ployed	In labor market training	Outside of the labor force	Missing	Total
Labor market status in the survey data						
Employed	262 (64.2 %)	10 (2.2 %)	1 (1.5 %)	8 (10.5 %)	0 (0.0 %)	281 (27.9 %)
Unemployed	64 (15.7 %)	376 (83.0 %)	7 (10.3 %)	14 (18.4 %)	2 (66.7 %)	463 (45.9 %)
In labor market training	11 (2.8 %)	4 (0.9 %)	54 (79.4 %)	7 (9.2 %)	0 (0.0 %)	76 (7.5 %)
Outside of the labor force	12 (2.9 %)	3 (0.7 %)	0 (0.0 %)	33 (43.4 %)	1 (33.3 %)	49 (4.9 %)
Missing	59 (14.5 %)	60 (13.3 %)	6 (8.82)	14 (18.4 %)	0 (0.0 %)	139 (13.8 %)
Total	408 (100.0 %)	453 (100.0 %)	68 (100.0 %)	76 (100.0 %)	3 (100.0 %)	1008 (100.0 %)

4. Results

In exploring job search training programs, we would like to know whether the treated, ($T=1$) benefited from treatment. One possible parameter for assessing average benefits of job search programs is the effect of treatment on the treated:

$$E[Y_1 - Y_0|T=1] = E[Y_1|T=1] - E[Y_0|T=1],$$

where Y_1 refers to the outcome if treated and Y_0 to the outcome without treatment. The fundamental problem here is that the counterfactual labor market outcome $E[Y_0|T=1]$ is unobserved. Random assignment solves this problem by randomly excluding some persons from being treated, creating a control group that is statistically similar with the treatment group in all respects except the treatment status.

In most cases evaluations are based on non-experimental methods where a suitably chosen comparison group substitutes data on the untreated outcome of the treated. The comparison by treatment status produces

$$E[Y_1|T=1] - E[Y_0|T=0] = E[Y_1 - Y_0|T=1] + \{E[Y_0|T=1] - E[Y_0|T=0]\}.$$

The last term presents selection bias that arises if the outcomes of non-treated are not representative of the outcomes of the treated had they not participated. Typically these pre-treatment differences are controlled for using regression or matching methods under the assumption of conditional independence (CIA). If this assumption holds $E[Y_0|X,T=1] = E[Y_0|X,T=0]$, and the non-experimental procedure yields similar results as the randomized experiment².

In the current context CIA can hold only if all relevant factors affecting participation in the experiment and subsequent employment are controlled. Given the data set at hand, we can balance different groups by sex, age, marital status, number and ages of children, broad occupation (10 categories), education (4 categories), local unemployment rate, whether the person lived in an urban area, NUTS3-level region, duration of the on-going unemployment spell, total months in unemployment during past three years and type of unemployment benefits received at the time of randomization. Following the example of Knight et al.

² Rosenbaum and Rubin (1983) showed that CIA also holds when the propensity score, $P(X)$, is used in conditioning.

(2006) we also included indicators of the ten most common quarterly sequences of past six labor market transitions.

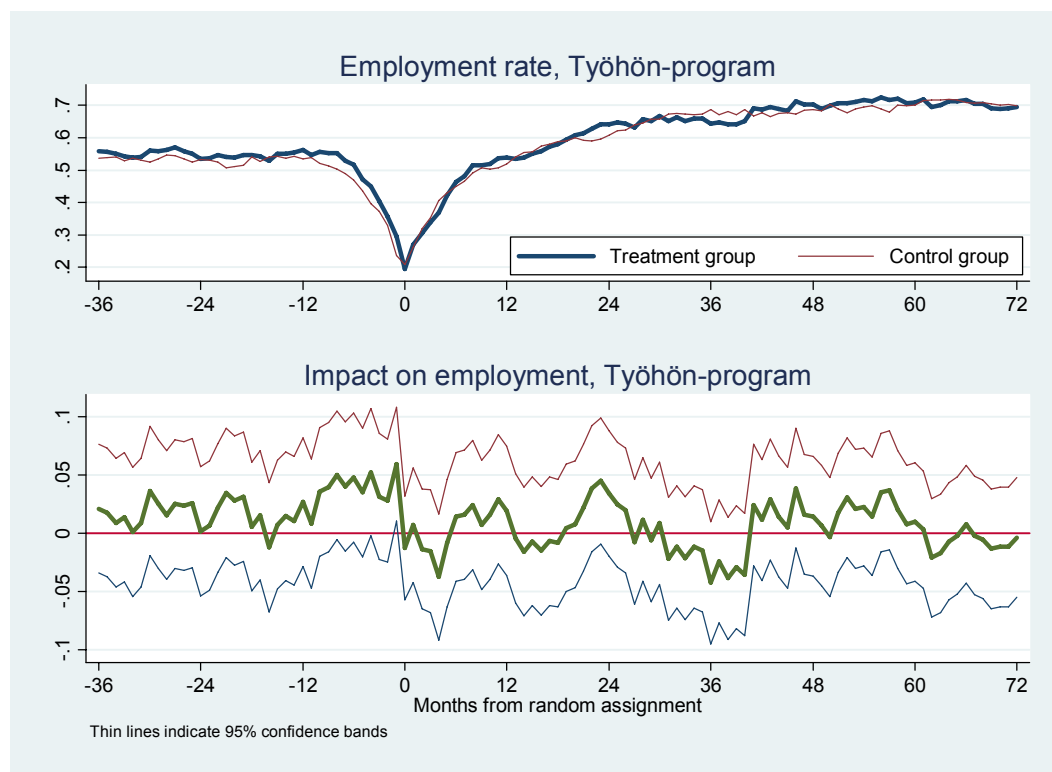
However, even with very rich data, some differences between the participants and non-participants may remain and cause bias in the estimates. Particularly differences in factors such as job-search intensity, health or motivation are hard to control for. In line with Heckman et al. (1998) we explore the role of unobserved factors by comparing the experimental control group to the matched non-experimental comparison group. Since neither of these groups took part in the training program the estimates reveal how large would be the bias if a non-experimental comparison group were used instead of an experimental control group to evaluate the program effects. This procedure also provides directly a significance test for the potential bias.

4.1 Long-term effects on job search training

Figure 1 reports the monthly employment rates in the treatment and in the control groups of the Työhön-program starting from three years before the experiment and extending up to six years after experiment. The treatment group also includes those who were assigned to treatment, but did not start the program, so the differences between the treatment and the comparison group reported in the lower panel of the figure refer to the effect of intention to treat. As shown in the figure, employment rates were very similar before the experiment indicating no problems in the randomization procedure. Employment rates decline during the 12 months before program participation. Such a decline is a common feature of the training programs targeted to those who are in a weakest position at the labor market when they participate in a training program.

Employment rates increase rapidly after the experiment but the increase is very similar in the treatment and in the control groups. There is no evidence of statistically significant employment effects during the entire six year follow-up period. Still, the difference between the treatment and the control groups is almost five percentage points 24 months after randomization, exactly at the time when the second follow-up survey was conducted. The results are well in line with a previous study by Vuori and Silvonen (2005) that used data from the two-year follow-up. Still, our conclusions regarding the overall benefits of the program are more pessimistic demonstrating the benefits of register data where the outcomes can easily be observed over a longer period instead of a single survey date.

Figure 1. *Employment effects of the Työhön -program*

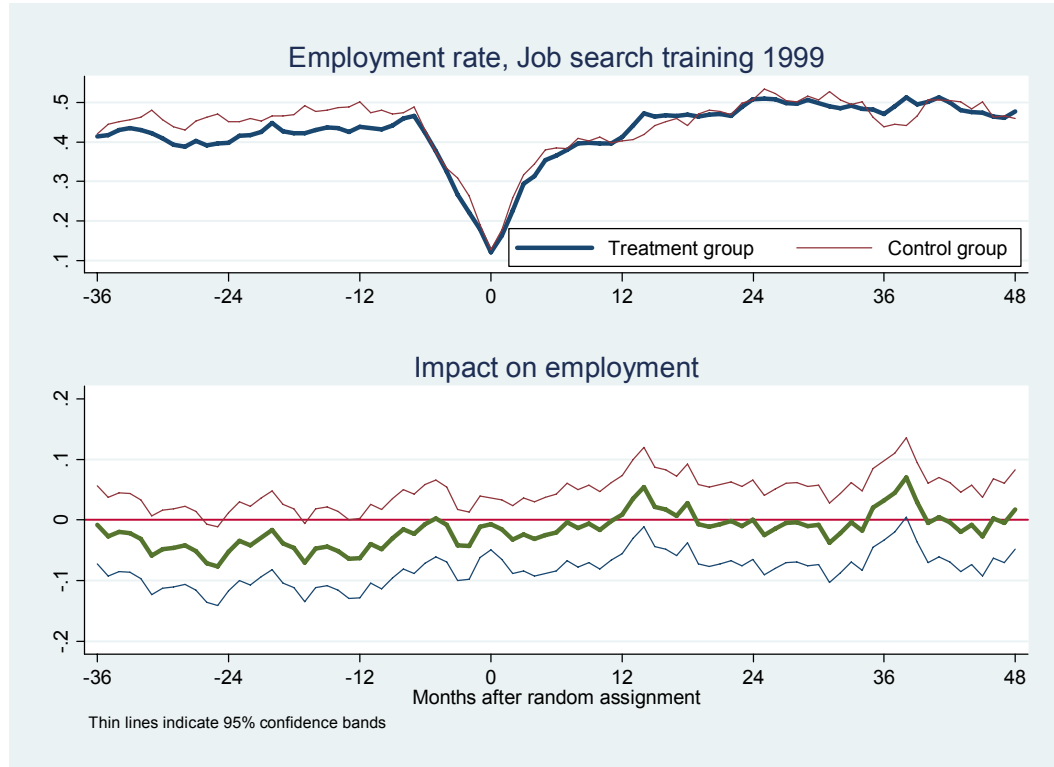


We repeat a similar analysis for the job search training experiment conducted in 1999. Figure 2 plots the employment rates in the treatment and the comparison groups from three years before to four years after the program. Again, the lower panel reports the differences in monthly employment rates and the confidence bands of these differences.

The results displayed in Figure 2 are generally similar to those in the first experiment. The difference in employment rates between the treatment and the control groups after the experiment is close to zero and generally not significant. However, even though treatment is randomly assigned there are some puzzling differences between the groups before the experiment. These differences are generally not statistically significant but we were still concerned that the experiment could have been somehow contaminated. We searched for possible explanations by checking the differences in pre-treatment employment rates separately in each employment office. We also suspected that caseworkers might have influenced the assignments but, since randomization was done at FIOH and the caseworkers only received a list of names of those who should be invited to training, this does not seem to be the case either. In the end we had to admit that we could not adequately explain the differences. However, even though this casts

doubts on the validity of experimental results it does not necessarily affect our comparison of experimental and non-experimental data.

Figure 2. *Employment effects of job search training 1999*



4.2 Comparison of experimental and non-experimental estimates

Before any comparisons between experimental controls and non-experimental comparison group, we transformed the comparison group data to person-month – observations and used simple logit-models to explain participation in the experiment. We then used propensity scores from the logit-model for balancing the observable characteristics of the experimental control and the non-experimental comparison group. In the Appendix A we report the average values of all covariates in the matched comparison group and compare these to the experimental control group averages. The Appendix A table is produced using Kernel matching but other matching methods seem to balance the distribution of the covariates almost equally well. In Kernel matching none of the 50 covariates included in the model differ in a statistically significant way between the experimental control and the non-experimental comparison groups. Since observing the comparison group from the same labor market has been shown to

be important in previous papers, we also experimented with matching procedures that limit the matched comparisons to those in the same region, but this had only a trivial effect on the results.

In Table 2 we compare the post-treatment employment rates in the experimental control group and in the matched comparison groups. The first column reports the employment rate in the experimental control group 6, 12, 24, 36, 48, 60 and 72 months after randomization. The next columns report the difference in employment rates between the various matched comparison groups and the experimental controls. These differences can be interpreted as a bias that would result if experimental control group were not available and the program effects were evaluated by comparing the treatment group to matched non-experimental comparison groups. In the second column we use simple nearest neighbor propensity score matching. In the third column we use five nearest neighbors which increases efficiency but may increase bias as matches become poorer. The columns 4 – 6 apply Kernel-matching that gives more weight to better matches. Column 4 uses all observations, Column 5 only observations in the region of “thick support” ($p < 0.15$), and Column 6 only persons that were unemployed at the date of randomization.

The results indicate that the bias resulting from the use of non-experimental matched comparison groups tend to be small if the effects are measured soon after program participation. Different matching methods tend to produce rather similar bias estimates. However, the bias tends to grow over time and becomes statistically significant two years after the experiment. At that point, the employment rate in the experimental control group is, on average, about five percentage points higher than the employment rate in the matched non-experimental comparison group. After three years the bias is about 10 percentage points depending on the matching method used.

Table 2. *Bias in the matching estimators for the employment effects of the Työhön program 1996/7*

Month	Experimental control group mean	Bias (experimental control - matched comparison)				
		Nearest neighbor	Five nearest neighbors	Kernel matching	Kernel matching 2	Kernel matching 3
6 th	0.449	0.013 (0.029)	0.007 (0.023)	0.010 (0.022)	-0.020 (0.023)	0.043 (0.026)
12 th	0.516	-0.006 (0.029)	-0.003 (0.023)	-0.001 (0.023)	-0.009 (0.023)	0.029 (0.027)
24 th	0.606	0.061 (0.029)	0.056 (0.023)	0.050 (0.022)	0.031 (0.023)	0.071 (0.027)
36 th	0.679	0.133 (0.028)	0.122 (0.022)	0.099 (0.021)	0.081 (0.022)	0.094 (0.026)
48 th	0.681	0.091 (0.028)	0.091 (0.022)	0.069 (0.021)	0.063 (0.022)	0.096 (0.026)
60 th	0.692	0.114 (0.028)	0.105 (0.022)	0.084 (0.021)	0.082 (0.022)	0.091 (0.026)
72 nd	0.689	0.151 (0.028)	0.124 (0.022)	0.112 (0.021)	0.091 (0.022)	0.125 (0.026)
N treated: on support		624	624	598	511	433
N treated: Total		624	624	624	512	464

Notes: Estimates are calculated using the psmatch2-procedure written for Stata by Leuven and Sianesi (2003). Common support restriction is imposed in the nearest neighbor estimates. Kernel estimates use epanechnikov kernel with bandwidth of 0.0004 selected using cross-validation among the register data comparison group. Experiments with alternative bandwidths produced similar results. Kernel 2 uses only observation in a region of thick support $p < 0.15$. Kernel 3 uses only register data on those unemployed during the first experiment. Standard errors are in parenthesis. Standard errors assume independent observations and do not count for estimated propensity score. Raw differences in employment rates range between -0.032 and 0.301. Further details on matching are in appendix available on request.

In Table 3, we calculate similar bias estimates using data from the 1999 job search training program experiment. The results turn out to be quite different in two respects. First, the employment rate in the experimental control group is much lower than in the first experiment. Second, none of the bias estimates are statistically significant indicating that matching methods could produce results that are similar to the results of randomized experiment in this case.

Table 3. *Bias in the matching estimators for the employment effects of the job search training programs 1999*

Month	Experimental control group mean	Bias (matched comparison - experimental control)				
		Nearest neighbor	Five nearest neighbors	Kernel matching	Kernel matching 2	Kernel matching 3
6 th	0.383	-0.057 (0.040)	-0.016 (0.031)	-0.032 (0.031)	-0.041 (0.032)	-0.001 (0.033)
12 th	0.404	0.009 (0.040)	-0.023 (0.031)	-0.011 (0.031)	0.005 (0.032)	0.055 (0.034)
24 th	0.509	0.024 (0.040)	0.031 (0.032)	0.019 (0.032)	0.023 (0.033)	0.057 (0.035)
36 th	0.437	0.006 (0.040)	-0.030 (0.032)	-0.009 (0.032)	0.001 (0.033)	0.026 (0.035)
48 th	0.464	0.012 (0.040)	0.001 (0.032)	0.025 (0.032)	0.008 (0.033)	0.051 (0.035)
N treated: on support		333	333	295	270	239
N treated: Total		335	335	335	278	274

Notes: Data ends in 2003, so we can only analyze the outcomes during four years after the experiment. The Kernel estimates on the rightmost column use only persons unemployed during the 1999 experiment. The cross-validated bandwidth is 0.0003. Raw differences in employment rates range between -0.195 and 0.169. Other notes under table 2.

There are several potential explanations for the differences in the results between the two experiments. For example, the unadjusted difference between the experimental control groups and the comparison group is initially much larger in the second experiment which may be reflected also in the differences between the matched groups. However, in both cases the matching procedure successfully balances the distribution of all observed characteristics.

Our favored interpretation of the difference in the bias estimates is that it is more difficult to control for the differences between the participants and the comparison groups when participation is entirely voluntary. Voluntary participation may lead to more selective group of participants (in terms of unobserved characteristics) than a case where program participation is determined by caseworkers. Since some important characteristics such as motivation are unlikely to be reliably measured in typical data used for program evaluation, this would also imply that non-experimental matching procedures produce more reliable results when program participation is based on more objective criteria and criteria that could be observed by the program evaluators.

5. Conclusion

Randomized experiments are the gold-standard of program evaluation. In addition to providing reliable estimates of the program effects they provide a benchmark against which to evaluate non-experimental evaluation methods.

In this paper we used data from two different randomized experiments with different selection processes of participants. The results are somewhat different. Typical matching methods yield reasonably good estimates when the selection to the program is based on the caseworker assessment. In contrast, even a large number of covariates do not get rid of the selectivity bias when program participation is entirely voluntary and perhaps largely determined by unobserved motivational or other characteristics of the applicants. In this case the resulting bias is substantial compared to the experimental estimate and could lead to quite different policy conclusions. Since selection processes differ across programs, results showing that non-experimental methods “work well” in a particular setting cannot be easily generalized to other programs.

References

- Abadie A. – Drukker D. – Herr J.L. – Imbens G.W. (2004): Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal* 4, 290-311.
- Abadie; A. – Imbens, G.W. (2006): Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica* 74, 235-268.
- Agodini, R. – Dynarski, M. (2004): Are experiments the only option? A look at dropout prevention programs. *The Review of Economics and Statistics* 86, 180-194.
- Bloom, H. – Orr, L. – Bell, S. – Cave, C. – Doolittle, F. – Lin, W. – Bos, J. (1997): The benefits and costs of JTPA title II-A programs: key findings from the national job training partnership act study. *The Journal of Human Resources* 32, 549-576.
- Blundell, R. – Costa Dias, M. – Meghir, C. – Van Reenen, J. (2004): Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association* 2, 569-606.
- Dolton, P. – O’Neill, D. (2002): The long-run effects of unemployment monitoring and work-search programs: experimental evidence from the United Kingdom. *Journal of Labor Economics* 20, 381-403.
- Heckman, J. – Ichimura, H. – Smith, J. – Todd, P. (1998): Characterizing selection bias using experimental data. *Econometrica* 66, 1017-1098.
- Hotz, V.J. – Imbens, G.W. – Klerman, J.A. (2006): Evaluating the differential effects of alternative welfare-to-work training components: a reanalysis of the California GAIN program. *Journal of Labor Economics* 24, 521-566.
- Knight, G. – Speckesser, S. – Smith, J. – Dolton, P. – Azevedo, J.P. (2006): Lone parents work focused interviews/New Deal for lone parents: combined evaluation and further net impacts. Department for Work and Pensions Research Report No 369.
- LaLonde, R. (1986): Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76, 604-620.
- Leuven, E. – Sianesi, B. (2003): PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. <http://ideas.repec.org/c/boc/bocode/s432001.html>.

- Malmberg-Heimonen, I. – Vuori, J. (2005): Activation or discouragement - the effect of enforced participation on the success of job-search training, *European Journal of Social Work* 8, 451 - 467.
- Michalopoulos, C. – Bloom, H. – Hill, C. (2004): Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *The Review of Economics and Statistics* 86, 156-179.
- Rosenbaum, P.R. – Rubin, D.B. (1983): The central role of the propensity score in observational studies of causal effects. *Biometrika* 70, 41-55.
- Van den Berg, G. – Van der Klaauw, B. (2006): Counseling and monitoring of unemployed workers: theory and evidence from a social experiment. *International Economic Review* 47, 895-936.
- Vinokur, A. – Price, R. – Schul, Y. (1995): Impact of the JOBS intervention on unemployed workers varying in risk for depression, *American Journal of Community Psychology* 23, 39-74.
- Vuori, J. – Silvonon, J. – (2005): The benefits of a preventive job search program on re-employment and mental health at 2-year follow-up. *Journal of Occupational and Organizational Psychology* 78, 43–52.
- Vuori, J. – Silvonon, J. – Vinokur, A. – Price, R. (2002): The Työhön job search program in Finland: Benefits for the unemployed with risk of depression or discouragement, *Journal of Occupational Health Psychology* 7, 5-19.

Appendix.

Appendix A. Average characteristics of the experimental control and matched comparison groups in the Työhön program 1996/7 and in Job training programs 1999. Kernel matching.

	Työhön 1996/7		Job search training 1999	
	Experimental control group mean	Non-experimental comparison group mean	Experimental control group mean	Non-experimental comparison group mean
Unemployed on the randomization date	0.74 0.73	0.52*** 0.74	0.82 0.79	0.40*** 0.79
Male	0.22 0.23	0.50*** 0.22	0.40 0.43	0.50*** 0.43
Age 25 – 34	0.37 0.37	0.25*** 0.35	0.29 0.29	0.26 0.30
Age 35 – 44	0.33 0.32	0.24*** 0.35	0.29 0.28	0.23*** 0.29
Age 45 – 54	0.17 0.18	0.21** 0.17	0.31 0.29	0.21*** 0.30
Single	0.40 0.40	0.45*** 0.39	0.44 0.46	0.47 0.44
Divorced or widow	0.13 0.13	0.14 0.14	0.16 0.17	0.15 0.18
Children, age 0-3	0.18 0.19	0.11*** 0.18	0.13 0.13	0.11 0.12
Children, age 4-6	0.09 0.09	0.06*** 0.09	0.08 0.08	0.08 0.08
Children, age 7-17			0.22 0.20	0.16*** 0.21
Disability	0.03 0.03	0.06*** 0.03	0.11 0.10	0.06*** 0.12
<i>Occupation</i>				
Unknown	0.04 0.04	0.09*** 0.04	0.07 0.06	0.08 0.06
Technical	0.08 0.09	0.06*** 0.08	0.06 0.06	0.07 0.06
Health care	0.06 0.07	0.07 0.07	0.13 0.13	0.09*** 0.12
Administrative	0.18 0.18	0.09*** 0.18	0.11 0.10	0.09 0.10
Trade	0.09 0.08	0.05*** 0.09	0.07 0.06	0.05 0.07

Industrial	0.09	0.17 ^{***}	0.15	0.17
	0.10	0.09	0.16	0.15
Services	0.10	0.09	0.16	0.09 ^{***}
	0.10	0.11	0.15	0.16
<i>Education (ref. primary)</i>				
Secondary	0.63	0.55 ^{***}	0.53	0.47 ^{**}
	0.63	0.64	0.53	0.53
Lower tertiary	0.09	0.06 ^{***}	0.18	0.18
	0.09	0.09	0.17	0.17
Higher tertiary	0.07	0.02 ^{***}	0.03	0.03
	0.07	0.06	0.03	0.03
<i>Month of participation</i>				
February			0.06	0.09 [*]
			0.07	0.07
March	0.07	0.09 ^{**}	0.20	0.09 ^{***}
	0.07	0.07	0.19	0.19
April	0.13	0.09 ^{***}	0.03	0.09 ^{***}
	0.12	0.13	0.03	0.03
May	0.17	0.09 ^{***}	0.04	0.09 ^{***}
	0.16	0.17	0.04	0.05
June	0.02	0.09 ^{***}	0.17	0.09 ^{***}
	0.02	0.02	0.18	0.18
July	0.09	0.09	0.20	0.09 ^{***}
	0.09	0.09	0.19	0.17
August	0.12	0.09 ^{***}	0.12	0.09
	0.12	0.12	0.13	0.12
September	0.10	0.09	0.00	0.09 ^{***}
	0.10	0.11	0.00	0.00
October	0.14	0.09 ^{***}	0.00	0.09 ^{***}
	0.14	0.13	0.00	0.00
November	0.04	0.09 ^{***}	0.02	0.09 ^{***}
	0.04	0.04	0.03	0.03
Lives in a city	0.81	0.59 ^{***}	0.64	0.61
	0.80	0.81	0.62	0.62
Lives in urban area	0.06	0.16 ^{***}	0.19	0.17
	0.06	0.06	0.21	0.20
Regional unemployment %	17.75	20.57 ^{***}	18.72	15.88 ^{***}
	17.77	17.92	18.36	18.24
Graduated in 1995/1998	0.04	0.05	0.05	0.04
	0.04	0.04	0.05	0.04
Received home care allowance in 1995/1998	0.10	0.05 ^{***}	0.07	0.03 ^{***}
	0.10	0.11	0.06	0.07
Months unemployed in past 3 years	12.36	14.98 ^{***}	14.91	13.14 ^{***}
	12.40	12.36	14.88	14.96

Duration of current unemployment, months	4.04	6.60 ^{***}	7.15	5.86 [*]
	4.08	4.16	7.27	7.34
Received labor market support in 1995/1998	0.14	0.23 ^{***}	0.33	0.31
	0.14	0.14	0.35	0.36
Received earnings related UI benefits in 1995/1998	0.41	0.41	0.46	0.37 ^{***}
	0.40	0.39	0.43	0.43
Received flat rate UA benefits in 1995/1998	0.12	0.18 ^{***}	0.04	0.04
	0.12	0.12	0.05	0.05
Does not seek full-time work	0.12	0.11	0.11	0.03 ^{***}
	0.12	0.12	0.08	0.09
<i>Prehistory</i>				
000000	0.28	0.28	0.18	0.39 ^{***}
	0.28	0.27	0.19	0.20
000001			0.02	0.03
			0.02	0.02
000011			0.01	0.02
			0.01	0.01
001111	0.01	0.03 [*]	0.02	0.02
	0.02	0.02	0.02	0.01
011111	0.01	0.02 [*]		
	0.01	0.01		
100000	0.09	0.04 ^{***}	0.10	0.04 ^{***}
	0.09	0.10	0.08	0.09
110000	0.06	0.03 ^{***}	0.09	0.03 ^{***}
	0.06	0.07	0.08	0.09
110011	0.03	0.02		
	0.03	0.03		
111000	0.04	0.03 [*]	0.04	0.02 ^{***}
	0.04	0.04	0.05	0.05
111100	0.04	0.04	0.05	0.02 ^{***}
	0.05	0.05	0.05	0.05
111110	0.02	0.03 ^{**}	0.05	0.02 ^{***}
	0.02	0.02	0.05	0.04
111111	0.05	0.18 ^{***}	0.12	0.13
	0.06	0.05	0.12	0.12
<i>Region</i>				
Varsinais-Suomi			0.00	0.09 ^{***}
			0.00	0.00
Satakunta			0.00	0.05 ^{***}
			0.00	0.00
Pirkanmaa			0.11	0.09
			0.13	0.12
Kaakkois-Suomi			0.01	0.06 ^{***}
			0.01	0.01

Etelä-Savo	0.10	0.04 ^{***}
	0.08	0.10
Pohjois-Savo	0.00	0.05 ^{***}
	0.00	0.00
Pohjois-Karjala	0.10	0.04 ^{***}
	0.11	0.09
Keski-Suomi	0.11	0.06 ^{***}
	0.12	0.12
Pohjanmaa	0.07	0.04 ^{***}
	0.07	0.08
Pohjois-Pohjanmaa	0.11	0.07 ^{***}
	0.13	0.14
Kainuu	0.16	0.02 ^{***}
	0.10	0.10
Lappi	0.09	0.05 ^{***}
	0.10	0.10

Notes:*** (**,*) = Difference between the groups is significant at the 1 (5, 10) per cent significance level. Upper figures correspond to unmatched mean values and lower figures to matched mean values.

Appendix B. Raw differences in employment rates between experimental controls and non-experimental comparisons

Table B.1. Työhön 1996/7 experiment

Month	Difference (experimental controls – non-experimental comparisons)		
	All observations	Thick support	Unemployed during the experiment
6 th	-0.032 (0.020)	-0.028 (0.022)	0.152 (0.020)
12 th	0.014 (0.020)	0.016 (0.022)	0.185 (0.021)
24 th	0.100 (0.020)	0.083 (0.022)	0.229 (0.022)
36 th	0.164 (0.020)	0.146 (0.022)	0.270 (0.023)
48 th	0.144 (0.020)	0.130 (0.022)	0.279 (0.023)
60 th	0.156 (0.020)	0.151 (0.022)	0.276 (0.023)
72 th	0.171 (0.020)	0.157 (0.022)	0.301 (0.023)

Table B.2. Job search training experiment 1999

Month	Difference (experimental controls – non-experimental comparisons)		
	All observations	Thick support	Unemployed during the experiment
6 th	-0.190 (0.027)	-0.195 (0.030)	0.105 (0.026)
12 th	-0.181 (0.027)	-0.177 (0.030)	0.088 (0.027)
24 th	-0.088 (0.027)	-0.100 (0.030)	0.169 (0.028)
36 th	-0.147 (0.027)	-0.126 (0.030)	0.095 (0.028)
48 th	-0.101 (0.027)	-0.094 (0.030)	0.119 (0.028)

Appendix C. Balancing properties- Densities of estimated propensity scores for unmatched groups

Figure C.1. Experimental controls

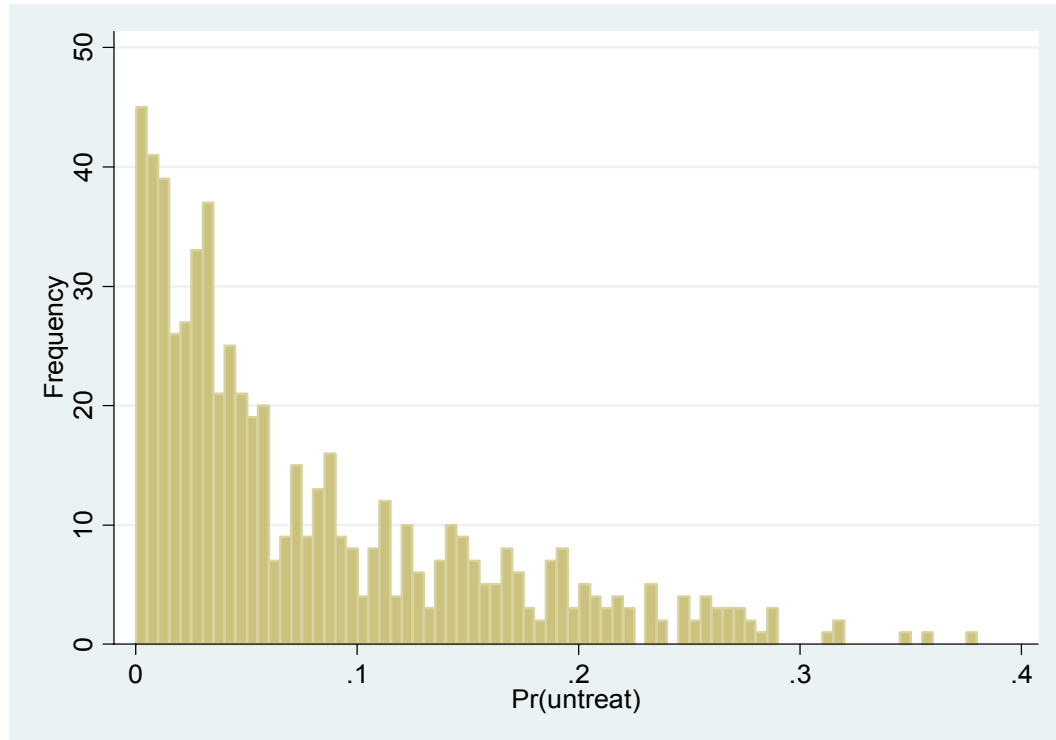


Figure C.2. Unmatched non-experimental comparison group: all observation

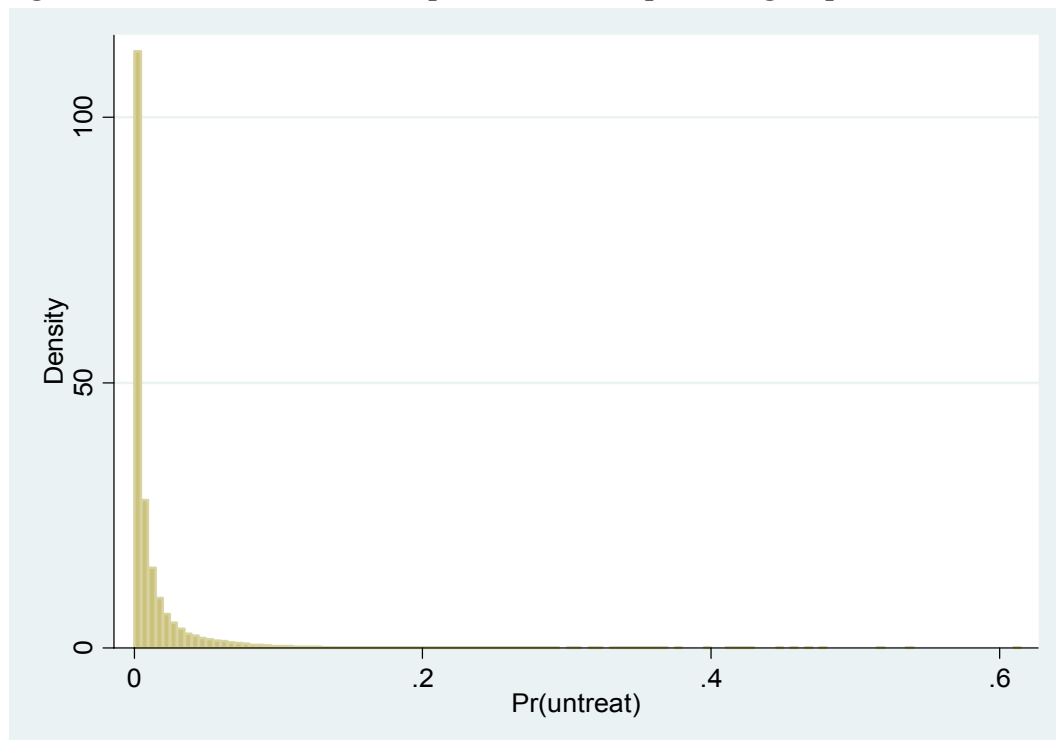
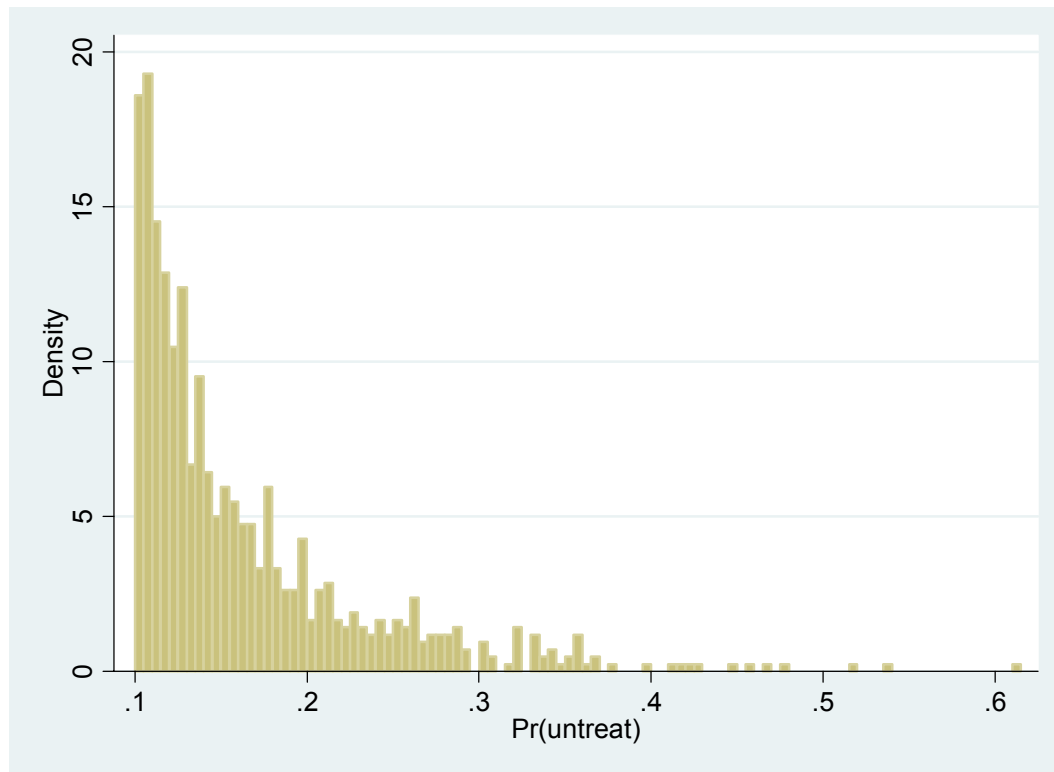


Figure C.3. Unmatched non-experimental comparison group: right-tale ($p > 0.1$)



Appendix D. Balancing properties - Distribution of estimated propensity scores for matched samples

Figure D.1. Työhön 1996/7 experiment – nearest neighbour (treated above)

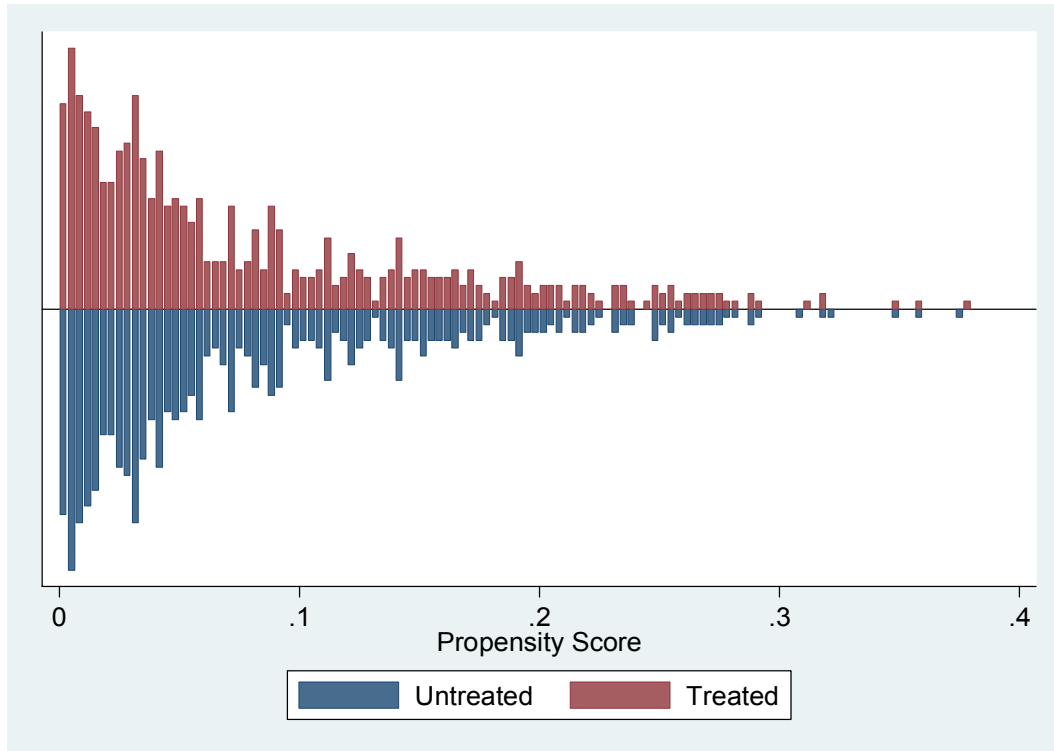


Figure D.2. Työhön 1996/7 experiment – 5 nearest neighbours (treated above)

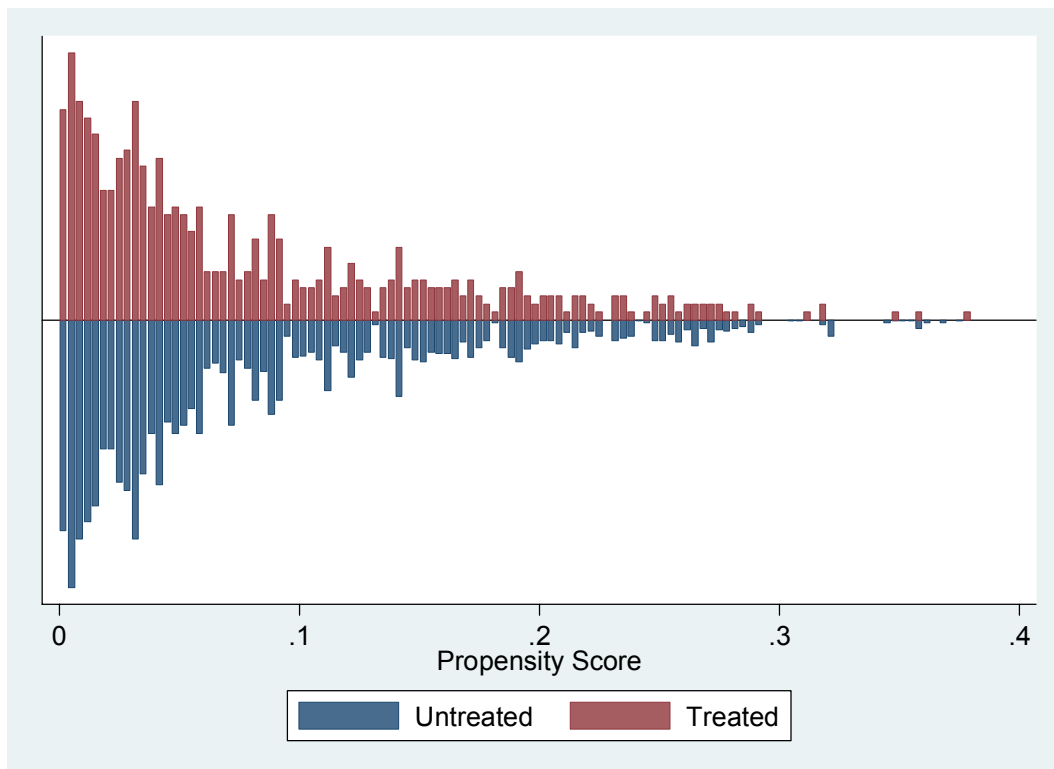


Figure D.3. Työhön 1996/7 experiment – kernel (treated above)

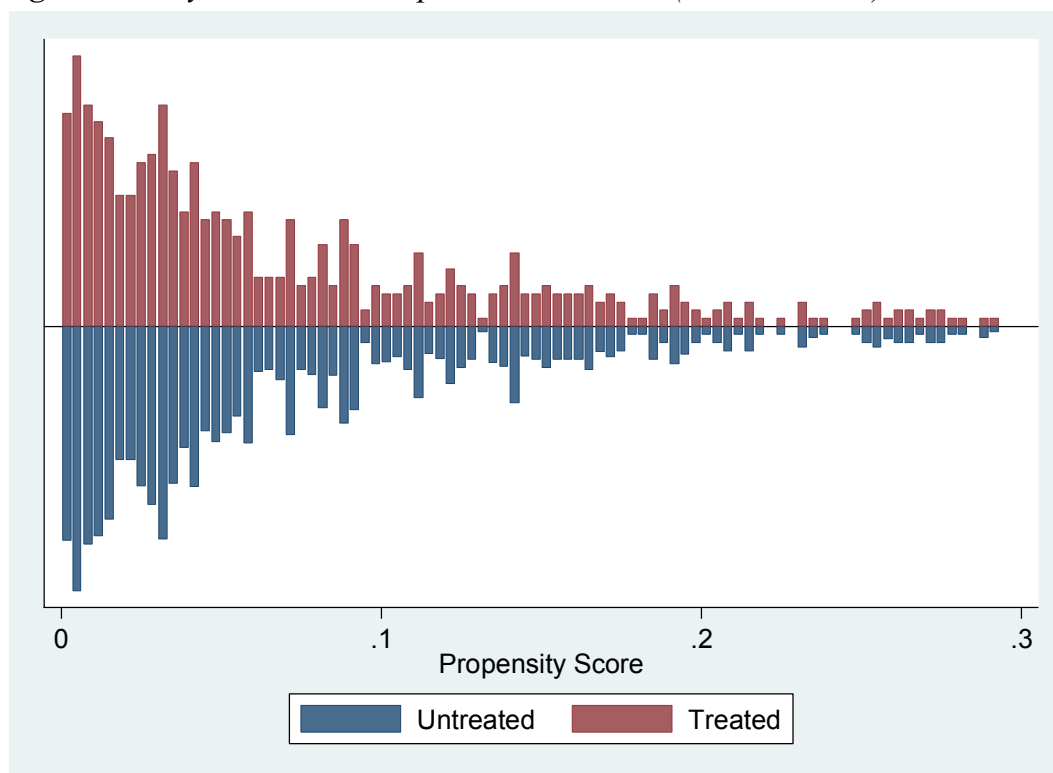


Figure D.4. Työhön 1996/7 experiment – kernel using thick support ($p < 0.15$, treated above)

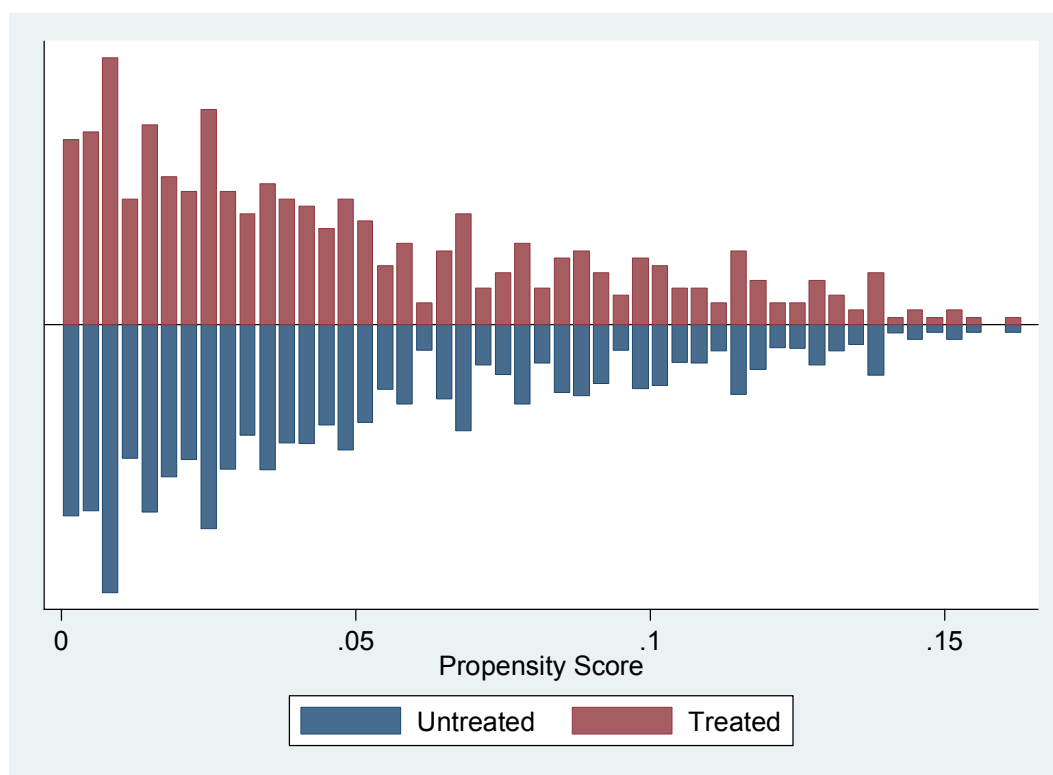


Figure D.5. Työhön 1996/7 experiment – kernel using unemployed only (treated above)

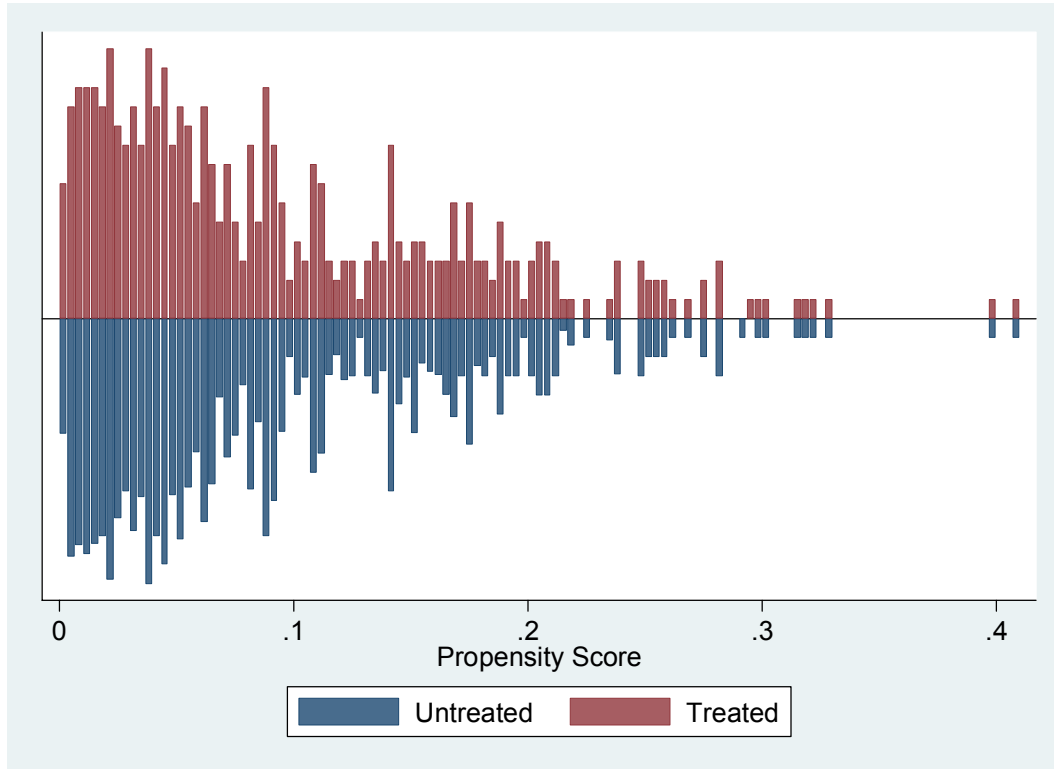


Figure D.6. Job search training experiment 1999 – nearest neighbour (treated above)

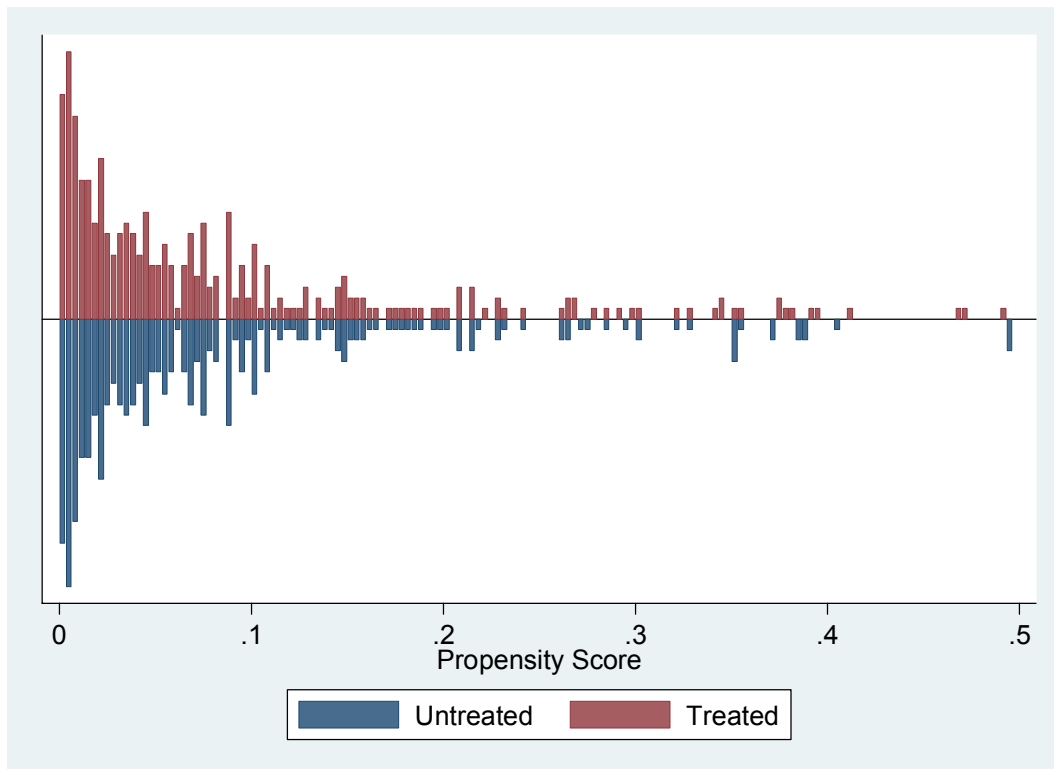


Figure D.7. Job search training experiment 1999 – 5 nearest neighbours (treated above)

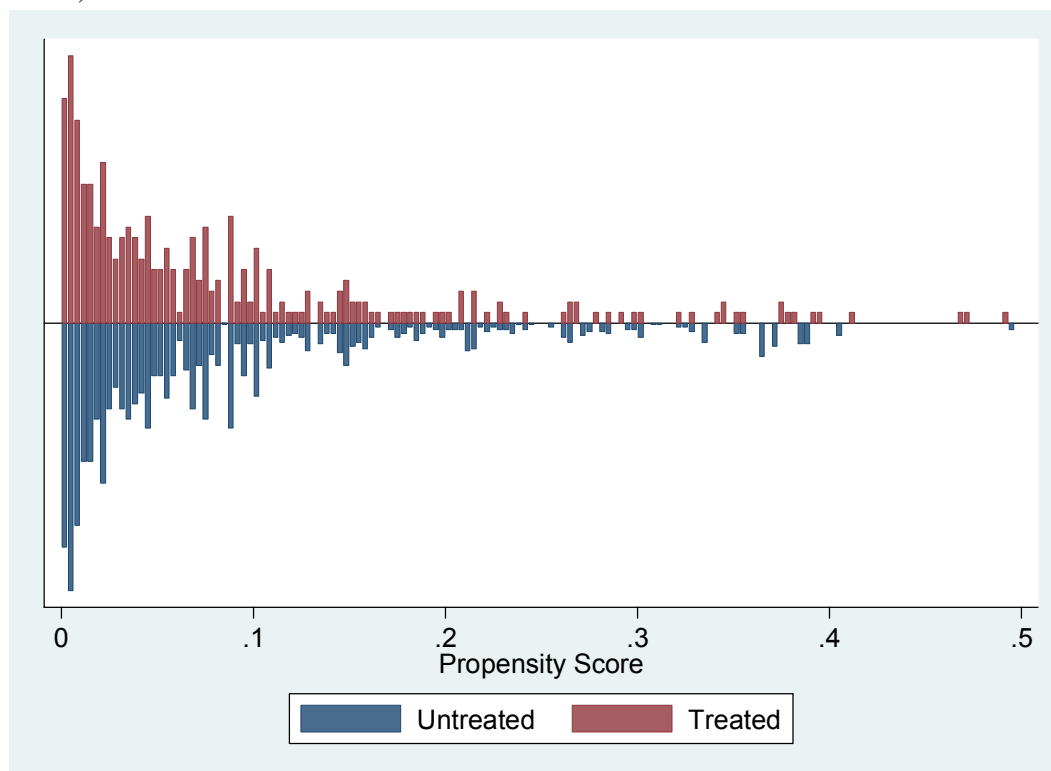


Figure D.8. Job search training experiment 1999 – kernel (treated above)

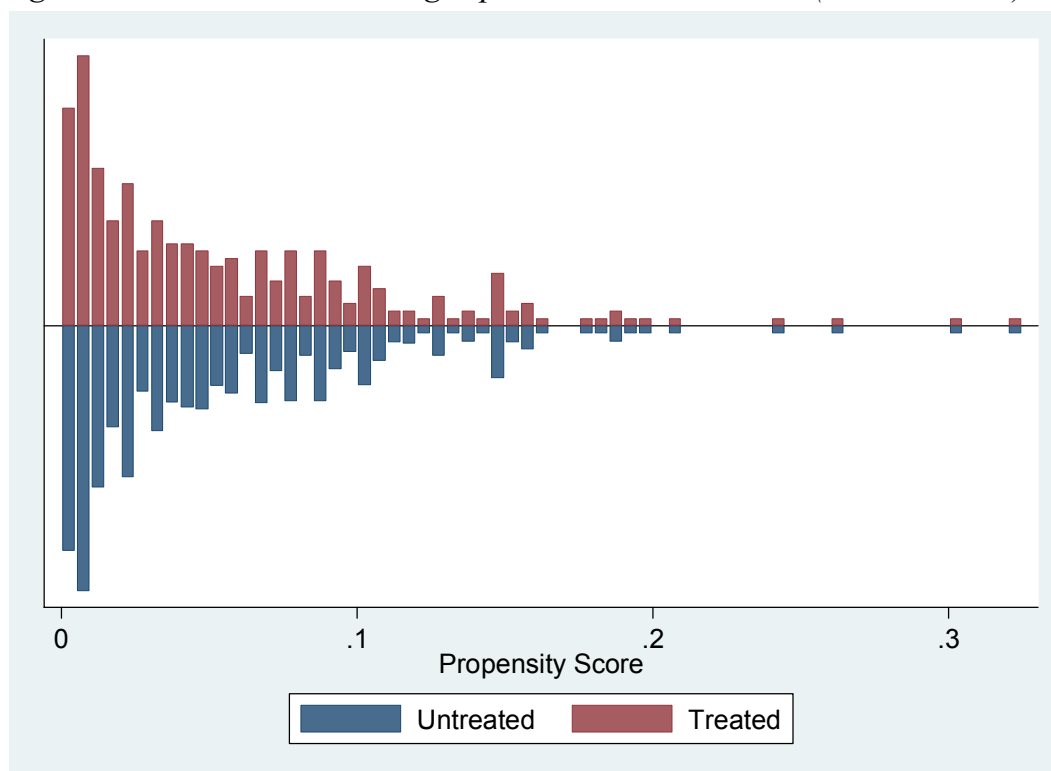


Figure D.9. Job search training experiment 1999 – kernel using thick support ($p < 0.15$, treated above)

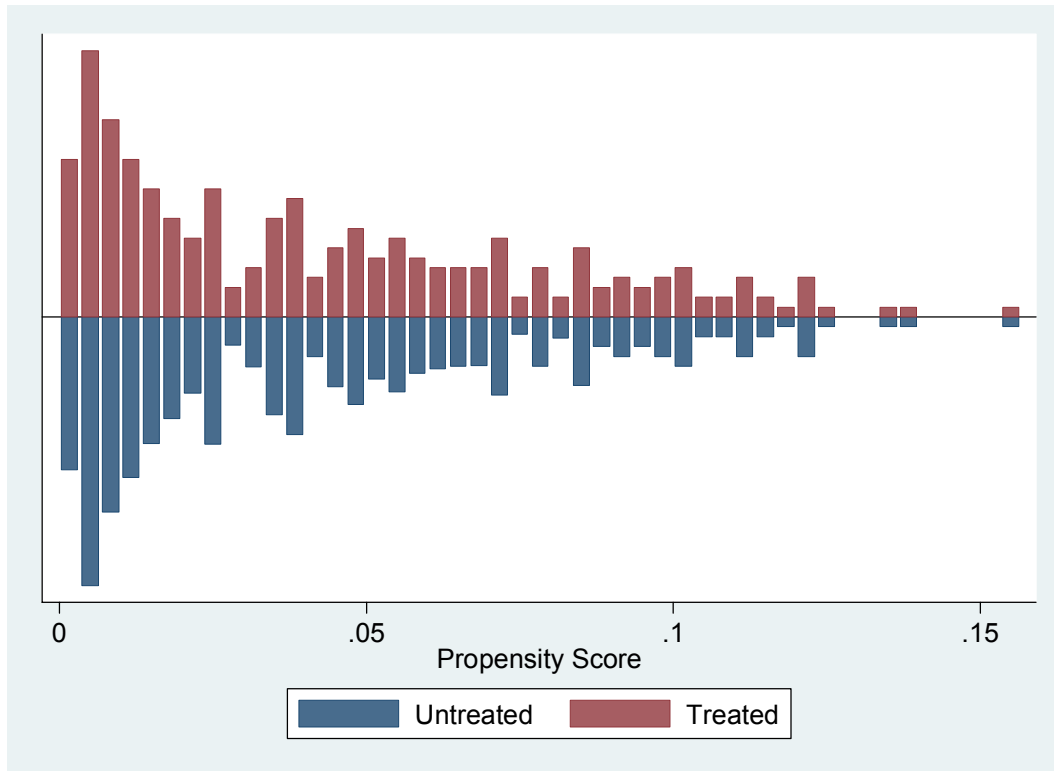
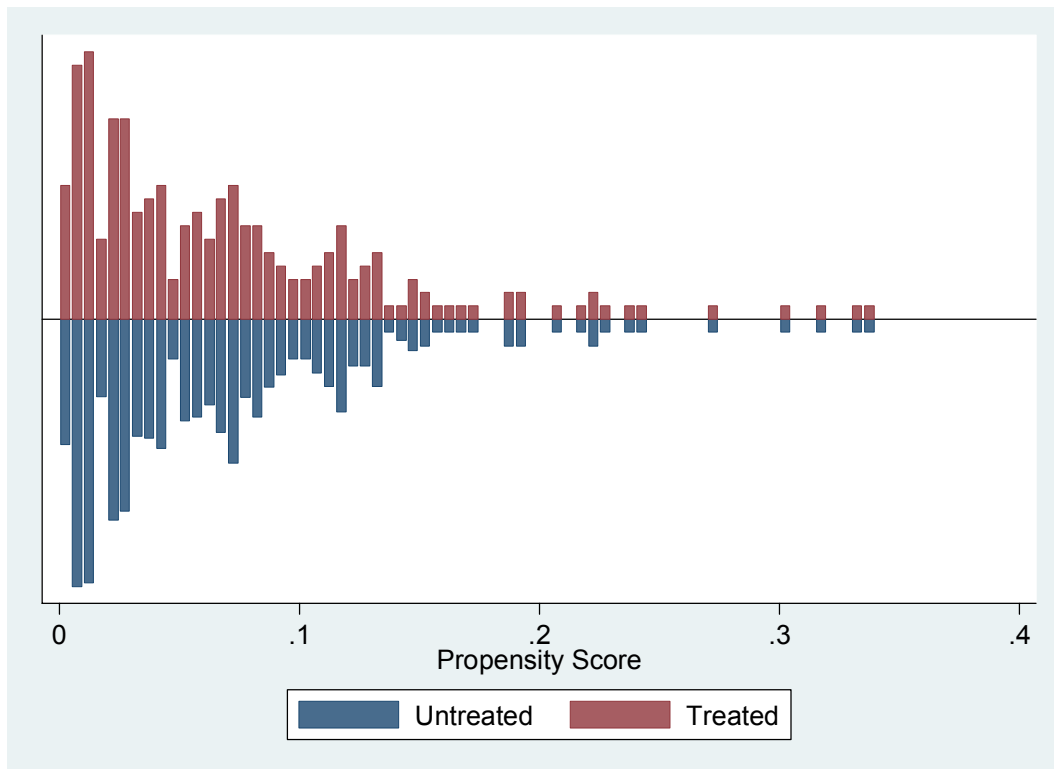


Figure D.10. Job search training experiment 1999 – kernel using unemployed only (treated above)



*Appendix E. Exact matching on region**Table E.1. The results of exact matching – Työhön experiment 1996/7*

Month	Original results		Exact matching		
	Experimental control group mean	Nearest neighbor	Five nearest neighbors	Nearest neighbor	Five nearest neighbors
6 th	0.449	0.013 (0.029)	0.007 (0.023)	-0.045 (0.032)	-0.039 (0.028)
12 th	0.516	-0.006 (0.029)	-0.003 (0.023)	-0.062 (0.031)	-0.058 (0.028)
24 th	0.606	0.061 (0.029)	0.056 (0.023)	0.033 (0.031)	0.026 (0.027)
36 th	0.679	0.133 (0.028)	0.122 (0.022)	0.097 (0.030)	0.096 (0.027)
48 th	0.681	0.091 (0.028)	0.091 (0.022)	0.044 (0.029)	0.051 (0.026)
60 th	0.692	0.114 (0.028)	0.105 (0.022)	0.093 (0.030)	0.107 (0.027)
72 nd	0.689	0.151 (0.028)	0.124 (0.022)	0.099 (0.030)	0.104 (0.027)

Notes: Estimations employ the `nmatch` module developed for STATA by Abadie et. al. (2004) Per cent of exact matches equals 100 in both cases. Standard errors use the variance formula in Abadie and Imbens (2006). This accounts for the uncertainty of the matching procedure but not the uncertainty involved in the estimation of unknown propensity score.

Table E.2. The results of exact matching – Job search training experiment 1999

Month	Original results		Exact matching		
	Experimental control group mean	Nearest neighbor	Five nearest neighbors	Nearest neighbor	Five nearest neighbors
6 th	0.383	-0.057 (0.040)	-0.016 (0.031)	-0.005 (0.037)	-0.000 (0.030)
12 th	0.404	0.009 (0.040)	-0.023 (0.031)	-0.008 (0.036)	-0.013 (0.031)
24 th	0.509	0.024 (0.040)	0.031 (0.032)	0.028 (0.038)	0.031 (0.032)
36 th	0.437	0.006 (0.040)	-0.030 (0.032)	0.020 (0.037)	0.038 (0.031)
48 th	0.464	0.012 (0.040)	0.001 (0.032)	0.029 (0.036)	0.049 (0.031)

Notes: Per cent of exact matches equals 100 in both cases. As in the results reported in text, three region dummies predict failure perfectly (regions in which the experiment did not take place) and those observations are dropped from analyses. Other comments as in the table above.

**VATT-KESKUSTELUALOITTEITA / DISCUSSION PAPERS ISSN 0788-5016
- SARJASSA ILMESTYNEITÄ**

381. Hjerppe Reino – Kiander Jaakko – Virén Matti: Are Government Expenditure Productive? Measuring the Effect on Private Sector Production. Helsinki 2006.
382. Riihelä Marja – Sullström Risto: Väestön ikääntyminen, kulutus, säästäminen ja eriarvoisuus. Helsinki 2006.
383. Hynninen Sanna-Mari – Kangasharju Aki – Pehkonen Jaakko: Regional Matching Frictions and Aggregate Unemployment. Helsinki 2006.
384. Ghatak Subrata – Sánchez-Fung José R.: Is Fiscal Policy Sustainable in Developing Economies? Helsinki 2006.
385. Lyytikäinen Teemu: Rent Control and Tenants' Welfare: the Effects of Deregulating Rental Markets in Finland. Helsinki 2006.
386. Riihelä Marja: Kotitalouksien kulutus ja säästäminen: Ikäprofiilien ja kohorttien kuvaus. Helsinki 2006.
387. Siivonen Erkki: Finanssisäännöt ja varallisuus oikeudet julkisten investointien analyysissa. Helsinki 2006.
388. Berghäll Elina: R&D and Productivity Growth in Finnish ICT Manufacturing. Helsinki 2006.
389. Berghäll Elina: Technical Efficiency in an R&D Intensive Industry: Finnish ICT Manufacturing. Helsinki 2006.
390. Berghäll Elina: Technical Change, Efficiency, Firm Size and Age in an R&D Intensive Sector. Helsinki 2006.
391. Ervasti Heikki – Venetoklis Takis: Unemployment and Subjective Well-being: Does Money Make a Difference? Helsinki 2006.
392. Hietala Harri – Kari Seppo: Investment Incentives in Closely Held Corporations and Finland's 2005 Tax Reform. Helsinki 2006.
393. Räisänen Heikki: Kaksi näkökulmaa julkisen työnvälityksen tehokkuuteen. Helsinki 2006.
394. Honkatukia Juha – Moilanen Paavo – Törmä Hannu: Runkoverkkosuunnitelman aluetaloudelliset vaikutukset. Helsinki 2006.
395. Honkatukia Juha – Rajala Rami – Sulamaa Pekka: Julkisen sektorin tuottavuuden kasvu ja työikäisen väestön määrän muutos 2005–2020, Rakenteellinen pitkän aikavälin tarkastelu alueellisella tasapainomallilla. Helsinki 2006.
396. Kyyrä Tomi – Wilke Ralf A.: Reduction in the Long-Term Unemployment of the Elderly: A Success Story from Finland Revised. Helsinki 2006.
397. Martikainen Emmi – Virén Matti: Valmisteverojen välittyminen kuluttajahintoihin Suomessa 1997–2004. Helsinki 2006.
398. Mälkönen Ville: Eri hankintamuodot julkisissa investoinneissa. Helsinki 2006.

399. Haataja Anita – Mattila-Wirola Päivi: Impact of Alternative Benefit Levels and Parental Choices on the Parents' Income. Micro-simulation Approach on the Finnish Parental Leave. Helsinki 2006.
400. Kyyrä Tomi – Ollikainen Virve: To Search or Not to Search? The Effects of UI Benefit Extension for the Elderly Unemployment. Helsinki 2006.
401. Hämäläinen Pellervo: Julkisten investointien tuottavuus. Katsaus kirjallisuuteen ja Suomi vuosina 1948-2003. Helsinki 2006.
402. Virén Matti: Fiscal Policy in the 1920s and 1930s. How Much Different It Is from the Post War Period's Policies. Helsinki 2006.
403. Aaltonen Juho: Perusterveydenhuollon menoeroja selittävät tekijät ja terveyskeskusten kustannustehottomuus. Helsinki 2006.
404. Venetoklis Takis: Guide to FUSSEP (Finnish University Students Socio-Economic Preferences) 2005 round. Helsinki 2006.
405. Honkatukia Juha – Mälkönen Ville – Perrels Adriaan: Impacts of the European Emission Trade System on Finnish Wholesale Electricity Prices. Helsinki 2006.
406. Kyyrä Tomi – Maliranta Mika: The Micro-Level Dynamics of Declining Labour Share: Lessons from the Finnish Great Leap. Helsinki 2006.
407. Korkeamäki Ossi – Uusitalo Roope: Employment Effects of a Payroll-Tax Cut: Evidence from a Regional Tax Exemption Experiment. Helsinki 2006.
408. Kari Seppo – Kiander Jaakko – Ulvinen Hanna: Vapaaehtoinen eläkevakuutus ja verotus. Katsaus kirjallisuuteen ja empiirinen kuva vapaaehtoisen eläkesäästämisen kehityksestä. Helsinki 2006.
409. Jalava Jukka – Kavonius Ilja Kristian: Durable Goods and Household Saving Ratios in the Euro Area. Helsinki 2006.
410. Sulamaa Pekka – Widgrén Mika: Turkish EU Membership: A Simulation Study on Economic Effects. Helsinki 2007.
411. Kohonen Anssi: Perintö- ja lahjaverotus – Näkökulmia talousteoriasta, maailmalta ja Suomesta. Helsinki 2007.
412. Perrels Adriaan: Economic Implications of Differences in Member State Regulations for the European Union Emission Trade System. Helsinki 2007.
413. Lehtonen Sanna – Moisio Antti: Kuntien valtionosuusjärjestelmä Suomessa ja Ruotsissa. Helsinki 2007.
414. Seppä Elina: Innovation Performance of Firms in Manufacturing Industry: Evidence from Belgium, Finland and Germany in 1998-2000. Helsinki 2007.
415. Kannianen Vesa – Kari Seppo – Ylä-Liedenpohja Jouko: Nordic Dual Income Taxation of Entrepreneurs. Helsinki 2007.
416. Kari Seppo – Karikallio Hanna: Tax Treatment of Dividends and Capital Gains and the Dividend Decision under Dual Income Tax. Helsinki 2007.
417. Perrels Adriaan – Kangas Elina: Vapaa-ajan asuntojen omistus ja käyttö – Esiselvitys ekotehokkuuden kartoitusta varten. Helsinki 2007.

418. Riihelä Marja – Sullström Risto – Tuomala Matti: Economic Poverty in Finland 1971–2004. Helsinki 2007.
419. Lyytikäinen Teemu: The Effect of Three-Rate Property Taxation on Housing Construction. Helsinki 2007.
420. Korkeamäki Ossi: Laskelmia miesten ja naisten välisen palkkaeron kaventamisesta julkisella sektorilla. Helsinki 2007.
421. Kosonen Tuomas: The Increased Revenue from Finnish Corporate Income Tax in the 1990s. Helsinki 2007.
422. Appelqvist, Jukka: Wage and Earnings Losses of Displaced Workers in Finland. Helsinki 2007.
423. Honkatukia Juha – Rajala Arto: Energia, päästökauppa ja kilpailukyky – Suomalaisen energiaintensiivisen teollisuuden näkemyksiä EU:n päästökaupasta ja pohjoismaisista energiamarkkinoista. Helsinki 2007.
424. Kari Seppo – Kosonen Tuomas – Kröger Outi: Vakuutusturvan vaje perhehuoltajan kuoleman kohdatessa. Julkisen turvan taso ja yksityinen henkivakuutusurva. Helsinki 2007.
425. Luoma Kalevi – Moisio Antti – Aaltonen Juh: Secessions of Municipal Health Centre Federations: Expenditure and Productivity Effects. Helsinki 2007.
426. Kari Seppo – Karikallio Hanna – Pirttilä Jukka: Anticipating Tax Changes: Evidence from the Finnish Corporate Income Tax Reform of 2005. Helsinki 2007.
427. Honkatukia Juha – Marttila Kimmo – Sulamaa Pekka: Budjetin aluevaikutukset – Valtion alueellistamis- ja tuottavuusohjelman vaikutukset maakunnissa. Helsinki 2007.
428. Kirjavainen Tanja: Efficiency of Finnish Upper Secondary Schools: An Application of Stochastic Frontier Analysis with Panel Data. Helsinki 2007.
429. Aaltonen Juh: Determinants of Health Care Expenditures in Finnish Hospital Districts 1993-2005. Helsinki 2007.
430. Haataja Anita: Soviteltu työttömyysetuus: Taustaa ja nykytilanne. Helsinki 2007.
431. Haataja Anita – Korkeamäki Ossi: Soviteltu työttömyysetuus: Kohdentuminen ja toimeentulo. Helsinki 2007.
432. Hämäläinen Kari – Tuomala Juha: Vocational Labour Market Training in Promoting Youth Employment. Helsinki 2007.
433. Parkkinen Pekka: Riittääkö työvoima terveydenhuolto- ja sosiaalipalveluihin? Helsinki 2007.
434. Kohonen Anssi: Yritysverotuksen koordinointi ja verokilpailu Euroopan unionissa. Helsinki 2007.
435. Berghäll Elina: Revealing Agglomeration Economies with Stochastic Frontier Modelling in the Finnish ICT Industry. Helsinki 2008.
436. Uimonen Sakari: Suomen infrastruktuuripääoma: Tiet. Helsinki 2007.
437. Lehtonen Sanna: Suomalaisten lukiodien tehokkuus – DEA yksilötason aineistolla. Helsinki 2007.