Quantitative Economics for the Evaluation of the European Policy

Dipartimento di Economia e Management

Co-funded by the Erasmus+ Programme of the European Union

Project funded by

European Commission Erasmus + Programme –Jean Monnet Action

Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Irene Brunetti Davide Fiaschi Angela Parenti[1]

9/10/2017

[1] ireneb@ec.unipi.it, davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

Lindley, J. et al. (2015), **A non-parametric evaluation of the Want2Work active labour market policy**

- Active labour market policies are popular tools used to help get unemployed people back into work.
- They are measures to improve the situation, in terms of employment and wages, of the unemployed, and of disadvantaged populations.
- In particular, they include: public employment services, labour market training, youth employment and training measures, subsidized employment, employment programs for the disabled, job rotation and job sharing, and direct job creation.

Introduction

- It is a common problem that interventions of this type are often established without an equivalent control group on which to base an evaluation of the program effectiveness.
- The Want2Work scheme is an active labour market policy that was introduced by the Welsh Assembly Government in order to improve the chances of individuals currently out of work re-entering the labour market.
- The Want2Work pilot scheme ran from September 2004 until March 2008 in particular areas of Wales.
- The primary aim of the scheme is to improve the re-employment chances of the participants

Introduction

- Want2Work was intended for Incapacity Benefits (IB) recipients.
- Thus, many of the key characteristics of the scheme were concerned with the health status of participants.
- The scheme was voluntary and advertisements were placed in public places such as doctors surgeries.
- The key features of the program include a combination of measures directed to improve the information of participants as well as to provide financial incentives.
- **Aim of the paper**: *all additional services, over and above standard assistance to those out of work, led to an increase in the likelihood that participants obtained a job?*

Problems

Two issues:

- A control group was not established as part of the original evaluation protocol;
- the participation to the program is voluntary.

$$\Downarrow$$

- Authors use **propensity score matchin**g techniques to derive a control group of non-participants with similar observed characteristics to those who participate in the program.
- Then, they compare the employment probabilities of each group.

N.B.: Propensity score matching does not impose any particular functional form on the estimated relationships.

The data

- They cannot observe participants in the non-participation state at the same point in time as they are participating.
- Data on non-participants must be used to estimate the counterfactual.
- Therefore, there must be good information available on both the treatment and control groups (participants and non-participants).

The data

- The database of information on Want2Work participants was collected by the Welsh Assembly Government.
- Any changes in status were also monitored and recorded.
- Who joined Want2Work between January 2005 and December 2007 is included $\Longrightarrow$ approximately **6,400 individuals** in the sample.
- **Detailed information on background characteristics:** age, gender, ethnicity, whether a single parent or not, highest qualification, type of welfare benefit being received when first registered with Want2Work, whether suffering from an illness or disability and if so what type, and **time spent out of work**.

The data

- **Time spent out of work** is a useful control variable: it is a proxy for unobserved employability characteristics.
- The **counterfactual data** used for the control group are drawn from the *Quarterly Labour Force Survey* (QLFS) for Great Britain.
- The QLFS has a wealth of information on employment outcomes and job characteristics, as well as all of the individual level characteristics that are observed for the Want2Work participants.

The data

- Want2Work data cover the period 2005 to 2007 $\Longrightarrow$ QLFS data for the same period as far as possible.
- Nine quarterly data sets for this period were used: March-May 2005, June-August 2005, and so on through to March-May 2007.
- All Want2Work participants were, by definition, initially out of employment $\Longrightarrow$ the QLFS sample was similarly restricted: excluding full time students and those who had taken early retirement.
- The control group was restricted to all those who responded to the survey for the full waves.
- The unemployment rate by travel-to-work area was therefore included amongst the list of conditioning variables.
- The resulting sample consisted of **8,994 men and women aged 16-65**. Of these, 3,427 reported that they wanted a job and were looking for a job.

The methodology

These data on the treatment and the control groups were used to estimate the effect of Want2Work:

$$ATET = E(Y_t|D = 1) - E(Y_c|D = 0) \tag{1}$$

It is assumed that the outcome for the control group provides a good estimate of the counterfactual for the treatment group.

Remind crucial assumption:

- **Conditional independence assumption**: $(Y_0; Y_1) \perp D|X$. Conditional on observed variables, X, the outcome is independent of treatment status.

The methodology

The propensity score, $e(x)$, is defined as the probability of an individual appearing in the treatment sample conditional on their observed characteristics:

$$e(x) = (Pr D = 1|X = x)$$

The propensity score can be estimated with a binary choice model such as a probit (or logit) equation.

Remind that:

- conditional on the propensity score, outcomes will be independent of treatment status: $(Y_0; Y_1) \perp D|e(x)$

Matching method

Two matching methods are performed:

- **Nearest neighbour matching**, and
- **Kernel matching**

- Kernel matching uses a weighted average of all of the observations in the control group to provide the match, with larger weights attached to observations with a closer propensity score to the treatment group individual being considered.
- The weights are inversely proportional to the distance between the scores.

A brief description of Kernel matching

Kernel matching estimator

$$ATET = \frac{1}{N^T}\{\sum_{j \in T} Y_i^T - \sum_{j \in C} w_{ij} Y_j^C\}$$

where w_{ij} are defined as:

$$w_{ij} = \frac{K(\frac{e_i - e_j}{h})}{\sum_{j \in C} K(\frac{e_i - e_j}{h})} \tag{2}$$

where K is the kernel density function and h is the chosen bandwidth.

Matching method

- The true propensity score is unknown and has to be estimated in the first stage of the procedure,
- The computation of the standard errors on the treatment effect, estimated in the second stage, needs to take this prior estimation into account.
- The usual approach is to bootstrap the standard errors.
- In particular, the procedure involves repeatedly calculating the treatment effect with random samples of the available data, to verify the degree of uncertainty attached to the result.

Why to use PSM?

- It does not impose any functional form on the relationship
- The technique identifies those observations in the treatment group for which there is no common support.

Results

Key indicator: individuals move into employment?

- Two indicators:
  - First indicator takes the value of one if individuals, in either the treatment or control group, moved into work at any point during the period in which they were observed.
  - A second indicator of labour market outcomes, for those who find a job, is the wage that they earn.
- Both full-time and part-time workers are considered.

Results: raw data on labour market outcomes

Table 1 √

Labour market outcomes for the treatment and control group.

a) Employment incidence (percent)

	Percentage that got a job	N
Want2Work	29.8%	6424
QLFS	18.7%	8994
QLFS seeking work	31.9%	3427

b) Annual earnings (percent in each pay band)

Gross annual pay	Want2Work participants	QLFS
<£10,000	54.7%*	61.8%
£10,000–14,999	35.4%*	16.7%
£15,000–19,999	7.7%	8.4%
£20,000–29,999	1.7%*	7.3%
£30,000+	0.5%*	5.8%
Mean	£9508†	£11,614
N	1901	838

Note: * denotes different from the QLFS at the 5% level. † estimated by an interval regression of the log of the wage band limits against a constant. QLFS is the GB Quarterly Labour Force Survey.

Propensity Score Matching Estimates of the Impact of Want2Work

Check the balancing property of PSM procedure.

- The matched sample will be balanced if there are no significant differences in the means of any characteristics between the treatment and control groups.
- They drop from the sample 2,415 observations that does not show tha common support.
- There remain some statistically significant differences in the characteristics between the two groups after the matching process.

Results: Propensity Score Matching Estimates of the Impact of Want2Work

Table 2
Propensity score estimates of W2W participation effect on probability of moving into work.

	Full sample		
	Propensity score (one to one)	Propensity score (kernel)	N
Any job	0.075** (0.018)	0.083** (0.015)	9299
Permanent job	0.109** (0.017)	0.104** (0.016)	9207
Full-time job	0.079** (0.016)	0.082** (0.010)	9299
	On Want2Work scheme < 15 months		
	Propensity score (one to one)	Propensity score (Kernel)	N
Any job†	0.065** (0.019)	0.073** (0.016)	9299
Permanent job	0.100** (0.017)	0.095** (0.015)	9207
Full-time job	0.074** (0.014)	0.077** (0.009)	9299
	IB recipients only		

Propensity Score Matching Estimates of the Impact of Want2Work

- Those who participated in Want2Work are 8 percentage points more likely to move into employment than similar job-seekers from the QLFS control group.
- This effect is both statistically and economically significant.
- Given that the average likelihood of moving into work in the sample is only around 30 percent, this impact of the Want2Work scheme is considerable.
- Two indicators of job quality that are available in both the Want2Work database and the QLFS: **whether or not the job acquired is full-time** and **whether or not the job acquired is permanent, or temporary and time-limited** in some way.

PSM Estimates of the Impact of Want2Work

Table 3
Propensity score estimates of W2W participation effect on earnings.

	Propensity score (one2one)	Propensity score (Kernel)	N
Full sample	−0.026	−0.003	2578
	(0.051)	(0.037)	
IB only	−0.043	−0.033	1134
	(0.144)	(0.085)	

Notes: matching variables are those listed in Table A2.

PSM Estimates of the Impact of Want2Work

- Estimates are provided for the full sample, and for IB claimants only.
- None of the four estimated coefficients approach statistical significance.
- All estimates are negative suggesting that Want2Work participants accept lower paid jobs, but the evidence is in no way conclusive.

Concluding Remarks

- The evidence produced is convincing, and supportive for Want2Work.
- The Want2Work participants consistently come out as being more likely to move into employment, compared to individuals in the control group.
- The size of the programme effect varies according to the specification being considered, but is always statistically significant.

Methodological note

- This paper shows how evaluations of labour market policies can be undertaken, even when policy-makers have not collected data on a control group of non-participants,
- Data from national surveys such as Labour Force Surveys can be used to obtain a sample of individuals in non-policy areas, who can then be matched to programme participants using propensity score matching techniques, to ensure the employment probabilities of similar individuals are being compared.
- Such non-parametric techniques have additional advantages that they do not impose functional form, and identify any individual for whom there is a lack of common support.

Applied Example in R

- Analyze the effect of going to Catholic school, as opposed to public school, on student achievement.
- Outcome variable: students' standardized math score ($c5r2mtsc_{std}$).
- Treatment variable: *catholic* (1 = student went to catholic school; 0 = student went to public school).

Tutorial in R

- Estimate the propensity score (the probability of being Treated given a set of pre-treatment covariates).
- Examine the region of common support.
- Choose and execute a matching algorithm: the nearest neighbor propensity score matching.
- Examine covariate balance after matching.
- Estimate treatment effects.

In R

library(MatchIt)
library(ggplot2)
$data <- read.csv("data.csv")$

Pre-treatment covariates

- $race_{white}$: Is the student white (1) or not (0)?
- $p5hmage$: Mothers age
- $w3income$: Family income
- $p5numpla$: Number of places the student has lived for at least 4 months
- $w3momed_{hsb}$: Is the mother's education level high-school or below (1) or some college or more (0)?

Pre-analysis using non-matched data

Before you implement a matching method, you'll conduct the following analyses using the non-matched data:

- Examine the difference-in-means between Treated and Control for the outcome variable.
- Examine the difference-in-means between Treated and Control for pre-treatment covariates.

Propensity score estimation

$m_{ps} < -glm(catholic \sim race_{white} + w3income + p5hmage + p5numpla + w3momed_{hsb}, family = binomial(), data = data)$

$summary(m_{ps})$

```
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.2125519  0.2379826 -13.499  < 2e-16 ***
## race_white   0.3145014  0.0700895   4.487 7.22e-06 ***
## w3income_1k  0.0073038  0.0006495  11.245  < 2e-16 ***
## p5hmage      0.0292168  0.0050771   5.755 8.69e-09 ***
## p5numpla    -0.1439392  0.0912255  -1.578    0.115
## w3momed_hsb -0.6935868  0.0743207  -9.332  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7701.3  on 9266  degrees of freedom
## Residual deviance: 7168.8  on 9261  degrees of freedom
##   (1811 observations deleted due to missingness)
## AIC: 7180.8
##
## Number of Fisher Scoring iterations: 5
```

Propensity score

Using this model, we can now calculate the propensity score for each student.

$prs_{df} < -data.frame(pr_{score} = predict(m_{ps}, type ='' response''), catholic = m_{ps}\$model\$catholic)$

$head(prs_{df})$

```
##     pr_score        catholic

## 1 0.2292928               0
## 2 0.1801360               0
## 4 0.2092957               0
## 5 0.2154022               1
## 6 0.3604931               0
## 7 0.1080608               0
```

Examining the region of common support

In R

$labs < -paste("\text{Actual school type attended:}", c(''Catholic'','' Public''))$

prs_{df}

$mutate(catholic = ifelse(catholic == 1, labs[1], labs[2]))$

$ggplot(aes(x = pr_{score}))+$

$geom_{histogram}(color ='' white'')+$

$facet_{wrap}(\sim catholic) + xlab("\text{Probability of going to Catholic school}")+$

$theme_{bw}()$

Executing a matching algorithm

- Restrict the sample to observations within the region of common support, and then to divide the sample within the region of common support into 5 quintiles, based on the estimated propensity score.
- Within each of these 5 quintiles, we can then estimate the mean difference in student achievement by treatment status.
- The method we use below is to find pairs of observations that have very similar propensity scores, but that differ in their treatment status. We use the package **MatchIt** for this.
- This package estimates the propensity score in the background and then matches observations based on the method of choice

$mod_{match} < -matchit(catholic \sim race_{white} + w3income + p5hmage + p5numpla + w3momed_{hsb}, method ='' nearest'', data = data)$

Executing a matching algorithm

To create a dataframe containing only the matched observations, use the *match.data*() function:

$data_m < -match.data(mod_{match})$

$dim(data_m)$

Examining covariate balance in the matched sample

- Visual inspection: plot the mean of each covariate against the estimated propensity score.
- t-tests of difference-in-means: t.test function.

Estimating treatment effects

$lm_{treat1} < -lm(c5r2mtsc_{std} \sim catholic, data = data_m)$

$summary(lm_{treat1})$

```
## lm(formula = c5r2mtsc_std ~ catholic, data = dta_m)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5089 -0.5754  0.0431  0.6167  3.0764
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.35728    0.02498  14.304  < 2e-16 ***
## catholic    -0.14762    0.03533  -4.179 3.02e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9185 on 2702 degrees of freedom
## Multiple R-squared:  0.006421,   Adjusted R-squared:  0.006054
## F-statistic: 17.46 on 1 and 2702 DF,  p-value: 3.024e-05
```