

Economic Policy Analysis (EC406)  
*Matching Estimators*

---

John Van Reenen  
*Centre for Economic Performance*  
*London School of Economics*

# The Program Evaluation Problem

---

---

Let  $d_i$  denote program participation of the  $i$ ' th individual

- $d_i = 1$  when individual  $i$  participates
- $d_i = 0$  when she does not

Denote individual  $i$ 's schooling by

- $S_{1i}$  when  $d_i = 1$
- $S_{0i}$  when  $d_i = 0$

The gain in schooling due to the program for an individual  $i$  who does in fact participate is

$$G_i = S_{1i} - S_{0i} \mid d_i = 1$$

Causal effect of program? **Average effect of the treatment on the treated:**

$$G = E(S_{1i} - S_{0i} \mid d_i = 1)$$

## The Program Evaluation Problem

---

---

**Evaluation problem:** do not typically observe the *counterfactual*,  $S_{0i}/d_i=1$   
(in this sense, program evaluation problem is a missing data problem)

We do observe the difference in mean outcomes between those who participate in the program and those who do not:

$$D = E(S_{1i} | d_i = 1) - E(S_{0i} | d_i = 0)$$

This is a biased estimate of average effect of the treatment on the treated

$$D = G + B$$

$$B = E(S_{0i} | d_i = 1) - E(S_{0i} | d_i = 0)$$

The bias is the expected difference in schooling in the absence of the program between those who participated and those who did not

## General Statement of the Problem (Homogeneous Treatment Effects)

---

---

### General specification of the outcome equation

$$Y_1 = g_1(X) + U_1$$

$$Y_0 = g_0(X) + U_0$$

- $X$  denotes observables
- $U$  denotes unobservables
- Allow different outcome functions for treatment and control groups

### Simple linear regression with a dummy variable for program participation corresponds to a special case where

- $E(U_1|X) = 0 = E(U_0|X) = E(U|X) = 0$
- $g_1(X) = X\beta_1$  and  $g_0(X) = X\beta_0$
- A more restrictive specification assumes all slope coefficients in  $\beta$  are the same across the two groups and only the intercept varies

## Average Effect of the Treatment on the Treated

---

---

The average effect of the treatment on the treated is

$$\alpha_T = E(Y_1 - Y_0 \mid X, d = 1)$$

Key assumption of the matching estimator is

- Conditional independence between non-treated outcomes and program participation (conditional on the observables  $X$ )

$$Y_0 \perp d \mid X$$

That is, *selection* only occurs on the *observables*

## The Matching Approach

---

---

Participation in the program may be non-random – individuals in the treatment group differ systematically from the population as a whole

However, under the assumption of *conditional independence*

- For each *treatment observation*,  $Y_1$ , look for a non-treated observation with the same realization of  $X$
- With selection only on observables,  $X$ , this non-treated observation constitutes the required counterfactual and becomes the *control observation*,  $Y_0$

Matching also assumes that  $0 < \text{Prob}(d = 1 / X) < 1$  in order to guarantee all treated agents have a counterpart in the non-treated population and that anyone constitutes a possible participant

- Existence of *common support*

## The General Form of the Matching Estimator

---

---

Let  $Z$  be the set of all possible values for the vector of explanatory variables,  $X$  (the support of  $X$ )

Let  $Z^*$  be the common support of  $X$  that is simultaneously observed among participants and non-participants in the sample

A consistent estimator for the average effect of the treatment on the treated,  $\alpha_T$ , is the empirical counterpart of

$$M(Z^*) = \frac{\int_{Z^*} E(Y_1 - Y_0 | X, d = 1) dF(X | d = 1)}{\int_{Z^*} dF(X | d = 1)}$$

That is, the mean difference in outcomes between treatment and control observations over the common support, weighted by the distribution of program participants

## The Use of the Propensity Score

---

---

When the observable characteristics,  $X$ , include a wide range of variables, matching can be very difficult due to the high dimensionality of the problem

Typically, more feasible to match on a function of  $X$ , in particular, on the **propensity score**

- On the probability of participating given the observable characteristics  $X$ ,  $P(X_i) = \text{Prob}(d_i = 1 | X_i)$

**THEOREM** : Rosenbaum and Rubin 1983, 1984

$$(Y_1, Y_0) \perp d | X$$

and

$$0 < \text{Prob}(d = 1 | X) < 1$$

$$\Rightarrow (Y_1, Y_0) \perp d | P(X)$$

Outcomes and program participation are independent conditional on the propensity score  $P(X)$

## Nearest Neighbor Propensity Score Matching

---

---

For each individual in the treatment group, find the non-treated individual with the most similar propensity score

The general form of the matching estimator simplifies to become

$$\hat{\alpha}_{MM} = \sum_{i \in T} (Y_i - Y_j) \frac{1}{N_T}$$

Where individual  $j$  is the nearest neighbor in terms of  $P(X)$  among the non-treated to individual  $i$  in the treatment group  $T$

(Note: the estimate of  $\alpha_T$  is based on comparing observations from the treatment group with observations from a control group which will typically be a subset of the sample of non-treated individuals)

# Recipe for Nearest Neighbor Propensity Score Matching

---

---

## Step 1

- Require representative samples of participants and of eligible non-participants

## Step 2

- Pool the two samples and estimate a logit model of program participation as a function of all variables ( $X$ ) in the data that are likely to determine participation

## Note: the matching estimator requires

- Assumption of conditional independence to be satisfied
- Assumption  $0 < \text{Prob}(d = 1 | X) < 1$  to be satisfied
- Model of program participation to have explanatory power

## Step 3

- Create the predicted values of the probability of participation from the logit regression. These are the propensity scores

# Recipe for Nearest Neighbor Propensity Score Matching

---

---

## Step 4

- Exclude non-participants with a propensity score outside the range observed for the treatment sample
- May wish to restrict potential matches in other ways (e.g. same geographical area to ensure matches from same economic environment)

## Step 5

- For each individual in treatment sample, find the observation from the non-participant sample with closest propensity score. This is its nearest neighbor

## Step 6

- The difference between the actual value for the treated observation and its nearest neighbor is the gain due to the program for that observation

## Step 7

- Calculate the mean of the individual gains for all individuals in the treatment group. This is the nearest neighbor propensity score matching estimate of the average effect of the treatment on the treated

## Nearest Neighbor Propensity Score Matching and Diff in Diff

---

---

The matching estimator can be implemented using cross-section data

However, with panel data before and after introduction of the program

- Can allow for an unobserved determinant of participation as long as it can be represented by separable individual and/or time-specific components of the error term

$$Y_{1it} = g_{1t}(X_{it}) + \eta_{1i} + \theta_t + \mu_{1it}$$

$$Y_{0it} = g_{0t}(X_{it}) + \eta_{0i} + \theta_t + \mu_{0it}$$

The nearest neighbor propensity score matching difference-in-differences estimator takes the following form

$$\hat{\alpha}_{MMDID} = \frac{1}{N_T} \sum_{i \in T} [(Y_{it_1} - Y_{it_0}) - (Y_{jt_1} - Y_{jt_0})]$$

## Empirical Application of Matching Methods

---

---

### Heckman, Ichimura and Todd, ReStud 1997, 605-54

- Examine data from a randomized experiment
  - National JTPA (Job Training Partnership Act) Experiment
    - Provides on-the-job training, job search assistance, and classroom training to youth and adults, who qualify for program under title IIA of the National Job Training Partnership Act
    - Persons become eligible for the program by having a family income near or below the poverty line for 6 months prior to application or by participating in federal, state or local welfare and food-stamp programs
    - Two-thirds of program applicants were assigned to treatment and one-third were randomized out and denied access to JTPA services for 18 months to form a randomized control group
    - Persons assigned to the control group only after they had applied to the JTPA program, been declared eligible, and had been accepted into the program

## Empirical Application of Matching Methods

---

---

The experiment yields randomized treatment and control groups

- Randomization as a solution to the program evaluation problem

Heckman et al. (1997) examine the empirical performance of matching methods by comparing the parameter estimates from randomization with those from non-experimental matching methods

- Consider a variety of non-experimental control groups
  - Eligible non-participants
    - Reside in the same narrow geographical region as participants and are eligible but do not apply
  - No shows
    - Experimental persons assigned to treatment who enrolled in JTPA but dropped out before receiving services
    - SIPP (Survey of Income and Program Participation) eligible
      - JTPA eligible, but may reside in different geographical location and not administered same questionnaire (survey instrument)
- Examine a variety of matching estimators

## Bias Decomposition

---

---

Consider the bias term,  $B$ , from the program evaluation problem

- Expected difference in outcomes in the absence of the program between the treated and the non-treated

This bias term can be decomposed into three components

$$B = E(Y_0 | X, d = 1) - E(Y_0 | X, d = 0) = B_1 + B_2 + B_3$$

- $B_1$  is the bias due to non-overlapping support of  $X$
- $B_2$  is the error due to misweighting on the common support of  $X$ , as the empirical distributions of treated and non-treated may not be the same even when restricted to the common support
- $B_3$  is the true econometric selection bias resulting from ‘selection on unobservables’

Matching methods try to correct for the first two sources of bias by choosing and reweighting observations

## Two Additional Sources of Bias

---

---

### Two additional potential sources of bias when using non-experimental methods

- Mismatch of the geographic location of program participants and non-treated individuals
- Mismatch of the survey instruments used to collect data on participants and non-treated individuals

# Empirical Findings

---

---

## Empirical findings of Heckman et al. (1997)

- Matching methods typically perform well in evaluating the impact of a major job training program
  - Impact estimates fairly close to those from a randomized evaluation
  - Importance of eliminating bias due to non-overlapping support and incorrect weighting of observations
  - Nonetheless, estimated selection bias (selection on unobservables) typically remains a substantial fraction of the experimentally-estimated program impacts
- Methods that perform most successfully involve
  - Combination of matching and difference in differences
  - Placing non-participants in the same labor market
  - Administering the same questionnaire
- The most quantitatively important sources of bias in evaluating the job training program with non-experimental methods were
  - Geographic mismatch and questionnaire mismatch
  - And not selection bias (selection on unobservables)

# Effects of a Mandatory Welfare to Work program: British New Deal for the Young Unemployed

Blundell, Costa-Dias, Meghir and Van Reenen (2004), JEEA

---

---

## Motivation

- What is the impact of welfare to work reforms (US and UK in 1980s and 1990s, now in Germany, France)? Aim to increase employment and reduce unemployment?
- In 1998 UK new government introduced the “New Deal for the Young Unemployed”
- After 6 months unemployment all young people aged 18-24 are enrolled into the program (mandatory)
- For 4 months they are intensively monitored and given job search assistance. If still no job then can get employer wage subsidy (or training/education)

# The Structure of the New Deal Program

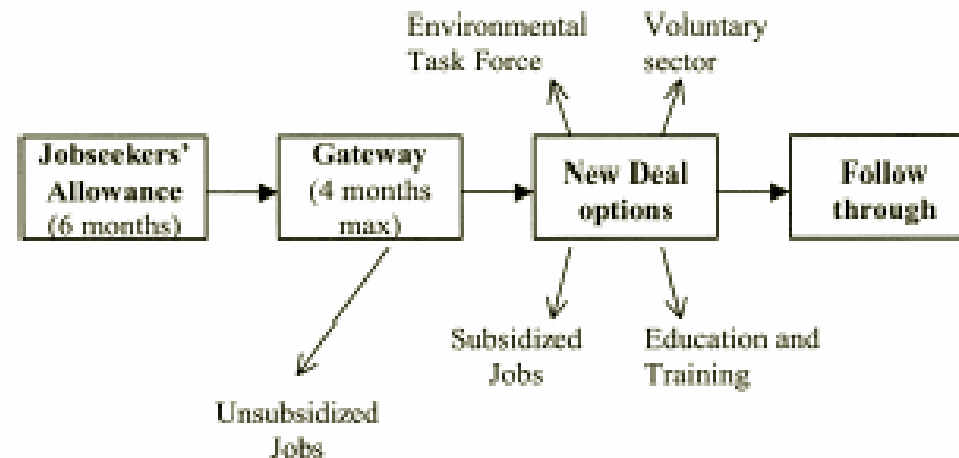


FIGURE 1. A simplified flow diagram of the New Deal program.

Outcome: probability of moving into a job during the 4 month Gateway period of job search (conditional on 6 months unemployment)

## BCMVR: Methods

---

---

- **Combined Difference in Differences with Matching**
- Difference in Differences has two possible comparison groups:
  - **Area.** New Deal was “piloted” in some areas for 3 months (Jan-March 1998) before it was introduced nationwide (April 1998). Can compare 19-24 year olds with 6 months unemployment in pilot areas (treatment) vs. 19-24 year olds with 6 months unemployment in non-pilot areas (comparison)
  - **Age.** 24 year olds and under were eligible, 25 year olds and older were not eligible. For example, 19-24 group in pilot areas vs. 25-30 year old group in pilot areas

## BCMVR: Methods

---

---

- Matching
  - Match areas using the pre-policy period trends in unemployment outflows (data back until 1982)
  - Choose areas whose trends most closely match those in the pilot areas

## Pre-policy Trends

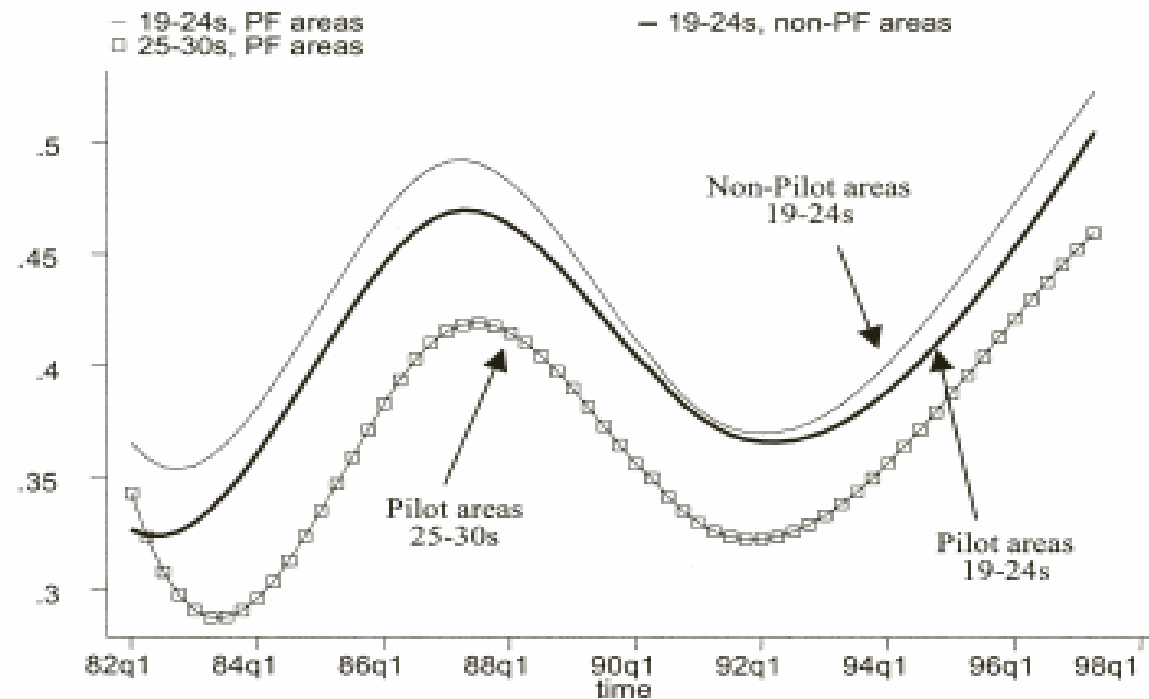


FIGURE 2. Outflows from JSA conditional on completing six months effect by the end of the 10th month on JSA. This graph illustrates the proportion of men leaving unemployment between the sixth and tenth months of unemployment 1982–1998. “PF” indicates that the men were living in a Pathfinder Pilot area (prior to New Deal introduction in 1998). The data have been smoothed by a cubic spline in time. Breakpoints were included at the first quarter of 1987 and the first quarter of 1990. No other covariates were considered.

Experiment	Treatment group	Comparison group	Number of observations	Linear Matching (OLS/Linear probability model)	Non-linear matching with non-additive error term (Logit specification)	Propensity score matching using smoothing splines	Non-linear matching using smoothing splines (Logit specification)
(1)	19-24 year olds living in Pathfinder areas	19-24 year olds living in all non-Pathfinder areas	3,716	0.110** (0.039)	0.098** (0.039)	0.104** (0.046) (0.024;0.182)	0.098** (0.044) (0.015;0.176)
(2)	19-24 year olds living in Pathfinder areas	19-24 year olds living in matched non-Pathfinder areas	1,193	0.134** (0.053)	0.073 (0.060)	0.093 (0.073) (-0.015;0.226)	0.080 (0.063) (-0.018;0.190)
(3)	19-24 year olds living in Pathfinder areas	25-30 year olds living in Pathfinder areas	1,096	0.104* (0.055)	0.091 (0.057)	0.078 (0.079) (-0.050;0.195)	0.074 (0.069) (-0.068;0.182)
(4)	19-24 year olds living in Pathfinder areas	31-40 year olds living in Pathfinder areas	1,169	0.159** (0.050)	0.096 (0.062)	0.099* (0.078) (-0.015;0.231)	0.082 (0.082) (-0.063;0.205)
<i>Outflow into the employment option (affecting 19-24 year olds in Pathfinder areas)</i>			4,486	0.057			
(5)	25-30 year olds living in Pathfinder areas	25-30 year olds living in all other areas	3,180	0.016 (0.042)	-0.012 (0.043)	0.027 (0.049) (-0.058;0.107)	0.031 (0.050) (-0.052;0.109)

## Findings

---

---

- Large impact of New Deal on outflows during pilot period (about 10 percentage points higher)
- Smaller (about 5 percentage points, but still significant) after National Roll Out in April 1998
- Robustness
  - How much is wage subsidy?
  - Effects before 6 months unemployment causing selection bias
  - Quality of jobs (no wages but look at jobs > 3months)
  - Women

## Conclusions

---

---

- Matching is a way of dealing with heterogeneous treatments
- Need rich set of covariates in  $X$
- Can be combined with difference in differences

## Conclusions

---

---

- Key issue in program evaluation is constructing the counterfactual – what is the right comparison group
  - True experiments
  - Heckman selection model
  - Instrumental Variables
  - Diff in Diffs
  - Matching
- In non-experimental data all about trying to get at the rule creating a good comparison group

# Essential Reading

---

---

## Essential reading

- Ravallion, M (2003) “The Mystery of the Vanishing Benefits: Ms Speedy Analyst’s Introduction to Evaluation”, World Bank, mimeograph
  - A condensed version of this paper is published as Chapter 3 in Baker, J (2000) *Evaluating the Impact of Development Projects on Poverty*, World Bank, Washington
- Blundell, R and Costa Dias, M (2000) “Evaluation Methods for Non-Experimental Data”, *Fiscal Studies*, 21(4), 427-68
- Heckman, J, Ichimura, H and Todd, P (1997) “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *Review of Economic Studies*, 64, 605-54