Econometrics II

Lecture 9
Matching

Marie Paul

University of Duisburg-Essen
Ruhr Graduate School in Economics

Summer Semester 2020

Outline of lecture

8. Matching
   - Potential outcomes framework
   - Identification
   - Estimation
   - Kernel Matching
   - Inverse Probability Weighting
   - Detailed Matching Application

Self-study plan for Lecture 9

- This lecture introduces matching methods.
- For our next (and last) virtual meeting, please work through all slides of Lecture 9. Readings:
- Cameron and Trivedi (2005) Chapter 25.4. and Angrist and Pischke (2009) Chapter 3.3. (maybe you want to read also earlier sections of chapter 3 if they are new to you) as far as this is covered on the slides.
- If you are interested in matching or dynamic treatment effects you may want to read Biewen et al. (2014) attached to this email.
- Please prepare answers to the "your own research" questions and collect questions and topics to discuss them in our virtual meeting.

Outline

8 Matching
- Potential outcomes framework
- Identification
- Estimation
- Kernel Matching
- Inverse Probability Weighting
- Detailed Matching Application

Motivation: Program evaluation

The **program evaluation literature** provides an intuitive statistical framework for the estimation of causal parameters.

It is commonly referred to as the **Neyman-Rubin causal model (RCM)** in recognition of **Rubin (1974, 1978)** and **Neyman (1923)**.

The RCM is centered around the notion of the distinction between **counterfactual** (or hypothetical) outcomes and **factual** (or realized) outcomes.

The fundamental problem is that causal inference requires observation of counterfactuals to answer "**what if**" questions.

At the heart of the RCM is the **potential outcomes model** from which the properties of many important estimators can be derived.

Motivation: Objects of interest

The program evaluation framework considers the effects of various **treatments**:

- Participation in a training programme on employment probability.
- Providing subsidies to disadvantaged regions.
- A university degree on earnings.

The ultimate aim is to **evaluate** these treatments to answer important **policy questions**:

1. Is the treatment **effective**? Is it **cost effective**?
2. What are the likely effects of implementing it in a **new context**?

Answering these questions require specifying a **counterfactual**:

- What are the **alternatives** to the treatment under scrutiny?
- What **would** the outcome have been in these cases?

Motivation

Example 8.1 (Your own research)

Have you ever used matching? Describe the application and the estimator used.

Potential outcomes

The main interest in the standard RCM is the **effect**, τ, of a **treatment**, $D = [0, 1]$, on an **outcome**, Y.

For a sample $i = 1, \ldots, N$, the triple (Y_i^1, Y_i^0, D_i) forms the basis of the **potential outcomes model (POM)** where

$$Y_i = \begin{cases} Y_i^1 & \text{if } D_i = 1, \\ Y_i^0 & \text{if } D_i = 0, \end{cases} \tag{8.1}$$

is the **observation rule** governing which of the **mutually exclusive** potential outcomes of Y_i we observe.

Note that

1. If subject i receives the treatment we observe Y_i^1, otherwise we observe Y_i^0 — but never **both simultaneously**
2. Equation (8.1) assumes that there are only **two** potential outcomes, but if there are, e.g., **peer effects** there could be many more.

SUTVA

For equation (8.1) to hold in practice we need to assume that **only** individual i:s treatment status matters for his or her own outcome. This implies that treatment does not **indirectly** affect outcomes of untreated.

This assumes away many potential mechanisms such as **peer effects** and **general equilibrium effects**.

Consider a binary treatment $D \in \{0,1\}$ with corresponding population vector $\mathbf{D} \in \{0,1\}^N$.

For any assignment vector $\mathbf{D}$, the outcome vector is $\mathbf{Y}(\mathbf{D})$ with individual outcome realizations $Y_i(\mathbf{D})$.

Assumption 8.1 (Stable Unit Treatment Value)

For each individual i and all possible pairs of treatment assignments $\mathbf{D}$ and $\mathbf{D}'$, we have

$$Y_i(\mathbf{D}) = Y_i(\mathbf{D}') \quad \text{if } D_i = D_i' \tag{8.2}$$

Treatment effects

If SUTVA holds, we can hence define two **potential outcomes** for each i as in equation (8.1).

The **evaluation problem** – i.e., that all relevant quantities are not observed – can then be written in the **switching** form

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0, \tag{8.3}$$

implying that we observe **one and only one** of two potential outcomes for each subject.

The **treatment effect** of D for individual i is then $Y_i^1 - Y_i^0$. The evaluation problem precludes us from identifying this effect.

The POM postulates a set of **average treatment effects** of interest:

- $\tau_{ATE} = \mathbb{E}\left[Y_i^1 - Y_i^0\right]$.
- $\tau_{ATT} = \mathbb{E}\left[Y_i^1 - Y_i^0 \mid D_i = 1\right]$.
- $\tau_{ATU} = \mathbb{E}\left[Y_i^1 - Y_i^0 \mid D_i = 0\right]$.

Outline

8. Matching
   - Potential outcomes framework
   - Identification
   - Estimation
   - Kernel Matching
   - Inverse Probability Weighting
   - Detailed Matching Application

The evaluation problem

Suppose we have a sample of smokers and non-smokers with information on **smoking status** (D_i) and each individual's subsequent **mortality** (Y_i).

How can we **identify** the effects of smoking? Suppose we just compare **averages**: mortality difference between smokers and non-smokers.

What **parameter** do we estimate? Note that

$$\begin{aligned}\mathbb{E}(\hat{\tau}) &= \mathbb{E}(Y_i^1 \mid D_i = 1) - \mathbb{E}\left(Y_i^0 \mid D_i = 0\right) \\ &= \mathbb{E}(Y_i^1 - Y_i^0 \mid D_i = 1) + \mathbb{E}\left(Y_i^0 \mid D_i = 1\right) - \mathbb{E}\left(Y_i^0 \mid D_i = 0\right) \\ &= \tau_{ATT} + \mathbb{E}\left(Y_i^0 \mid D_i = 1\right) - \mathbb{E}\left(Y_i^0 \mid D_i = 0\right).\end{aligned} \tag{8.4}$$

Thus, only if

$$\mathbb{E}\left(Y_i^0 \mid D_i = 1\right) - \mathbb{E}\left(Y_i^0 \mid D_i = 0\right) = 0 \tag{8.5}$$

our comparison gives us the **causal effect of smoking** (for smokers).

The selection problem

Equation (8.5) is of special interest. What is the implication of assuming that

$$\mathbb{E}\left(Y_i^0 \mid D_i = 1\right) = \mathbb{E}\left(Y_i^0 \mid D_i = 0\right)? \tag{8.6}$$

It means individuals may not **sort** into **treatment** due to factors related to the potential outcomes (here, Y^0). Formally,

$$(Y_i^1, Y_i^0) \perp D_i \quad \forall\, i = 1, \ldots, N. \tag{8.7}$$

Hence, in the smoking example it will fail whenever

- People **know** how smoking affects their mortality, and act upon it.
- People who **notice** adverse health effects **decide** to quit.
- People in worse/better **health** are more/less likely to smoke.
- People in worse/better health have different **risk preferences**.
- etc..

All these factors are clearly important so our simple comparison may **not** be given a causal interpretation.

Randomization: The gold standard

Suppose we take a **random sample** of n from the population of interest and **randomly assign** smoking status.

We can now estimate the treatment effect as

$$\hat{\tau} = \frac{1}{n_t} \sum_{i \in \Omega_T} Y_i - \frac{1}{n_c} \sum_{i \in \Omega_C} Y_i. \tag{8.8}$$

where, in a large sample, $\hat{\tau} \to \tau = \mathbb{E}\left(Y^1 \mid D = 1\right) - \mathbb{E}\left(Y^0 \mid D = 0\right)$.

Thanks to the **randomisation**, $\mathbb{E}\left(Y^k \mid D = k\right) = \mathbb{E}\left(Y^k\right) \forall k$. Thus,

$$\begin{aligned} \tau = \underbrace{\mathbb{E}\left(Y^1 | D = 1\right) - \mathbb{E}\left(Y^0 | D = 1\right)}_{ATT} &= \underbrace{\mathbb{E}\left(Y^1 | D = 0\right) - \mathbb{E}\left(Y^0 | D = 0\right)}_{ATU} \\ &= \underbrace{\mathbb{E}\left(Y^1\right) - \mathbb{E}\left(Y^0\right)}_{ATE}. \end{aligned} \tag{8.9}$$

Thus, controlling the **assignment mechanism** by randomizing it across subjects allows us to estimate the causal impact of smoking.

Randomization and regression

We may also derive $\hat{\tau}$ in a regression context:

$$Y_i = \alpha + \tau D_i + \epsilon_i. \tag{8.10}$$

Defining $X = (\mathbf{i}, \mathbf{D})$ and $\delta = (\alpha, \tau)'$, we get

$$\begin{aligned} \left(X'X\right)^{-1} &= \frac{1}{N_T\left(1-P\right)} \begin{bmatrix} P & -P \\ -P & 1 \end{bmatrix} \\ \hat{\delta}_{OLS} &= \left(X'X\right)^{-1} X'Y = \begin{bmatrix} \frac{1}{n_c} \sum_{i \in \Omega_C} Y_i \\ \frac{1}{n_t} \sum_{i \in \Omega_T} Y_i - \frac{1}{n_c} \sum_{i \in \Omega_C} Y_i \end{bmatrix} \end{aligned} \tag{8.11}$$

where $\mathbf{i}$ is a N vector of ones and P is the proportion treated.

When an experiment is correctly designed and implemented, it provides an **unbiased** estimate of the impact of a treatment.

Conditional Treatment Effects

What if we do not have an experiment but we can **control** for covariates X such as age, sex and ethnicity?

Divide the outcome measure into an **explained** and an **unexplained** part according to the **Roy model**:

$$Y = \begin{cases} Y^1 = \varphi_1(X) + U_1 & \text{if } D = 1, \\ Y^0 = \varphi_0(X) + U_0 & \text{if } D = 0, \end{cases} \tag{8.12}$$

where $\varphi_j(X) \equiv \mathbb{E}[Y^j|X]$ is **expected mortality** of individuals with characteristics X and smoking status j and U_j are **unobserved factors**.

Then the observed outcome Y from the switching model is then

$$Y = \varphi_0(X) + D\left(\varphi_1(X) - \varphi_0(X) + U_1 - U_0\right) + U_0, \tag{8.13}$$

where the second term is the effect of participation in the treatment.

Conditional Treatment Effects

With covariates, the effect of, e.g., smoking status on mortality is equal to

$$\tau(X) = \underbrace{\varphi_1(X) - \varphi_0(X)}_{\text{ATE}} + \underbrace{U_1 - U_0}_{\text{Selection bias}}. \tag{8.14}$$

If individuals unobservably **self-select** into treatment, this implies

$$\begin{aligned} &\mathbb{E}(U_0|X,D) \neq 0;\ \mathbb{E}(U_1|X,D) \neq 0 \\ &\mathbb{E}(U_1 - U_0|X,D) \neq 0, \end{aligned} \tag{8.15}$$

which leads to a difference between what we **estimate**

$$\mathbb{E}(\hat{\tau}(X)) = \varphi_1(X) - \varphi_0(X) + \mathbb{E}(U_1|X,D=1) - \mathbb{E}(U_0|X,D=0), \tag{8.16}$$

and our **parameters of interest**

$$\begin{aligned} \tau_{ATE}(X) &= \varphi_1(X) - \varphi_0(X), \\ \tau_{ATT}(X) &= \varphi_1(X) - \varphi_0(X) + \mathbb{E}(U_1 - U_0|X,D=1), \\ \tau_{ATU}(X) &= \varphi_1(X) - \varphi_0(X) + \mathbb{E}(U_1 - U_0|X,D=0). \end{aligned} \tag{8.17}$$

Selection on (un)observables

The critical assumption we need to make in order to identify the treatment effects in (8.17) is the **conditional independence (CIA) assumption**.

Assumption 8.2 (Conditional independence)

Conditional on X, the potential outcomes (Y^1, Y^0) are independent of treatment

$$(Y^1, Y^0) \perp D|X \tag{8.18}$$

The SUTVA and CIA jointly **identifies** all terms in (8.17).

However, to identify, e.g., the ATT effect, it is only necessary to assume

$$Y^0 \perp D|X, \tag{8.19}$$

or even less restrictive, the **conditional mean independence assumption**

$$\mathbb{E}\left(Y^0|X, D=1\right) = \mathbb{E}\left(Y^0|X, D=0\right) = \mathbb{E}[Y^0|X]. \tag{8.20}$$

Note that all these are assumptions of **selection on observables** meaning that we do not allow for **unobserved heterogeneity** once we have controlled for X.

General Setup: Assumptions

So far, the analysis has been evaluated conditional on a specific value of X.

In practice, we have to rely on an additional assumption of **common support** for identification of unconditional average treatment effects.

Assumption 8.3 (Common support)

Assume that for each realization of X there is at least some **overlap** *between treated and untreated subjects:*

$$0 < \Pr(D_i = 1 | X_i = x) < 1 \quad \forall x \tag{8.21}$$

The CIA or **unconfoundedness** assumption in (8.18) and the **common support** assumption are jointly denoted **strong ignorability**.

Definition 8.1 (Strong ignorability)

Assume that unconfoundedness (8.18) and common support (8.21) holds. This is defined as **strong ignorability**.

General setup: Identification

Strong ignorability allows us to estimate various causal effects:

$$\begin{aligned}\tau\left(x\right) =&\mathbb{E}\left[Y_i^1 - Y_i^0 | X_i = x\right] \\ =&\mathbb{E}\left[Y_i^1 | D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i^0 | D_i = 0, X_i = x\right] \\ =&\mathbb{E}\left[Y_i | D_i = 1, X_i\right] - \mathbb{E}\left[Y_i | D_i = 0, X_i\right].\end{aligned} \tag{8.22}$$

Given that we can identify $\tau\left(x\right)$ for all x, we can identify the treatment effect for different **populations**:

$$\begin{aligned}\tau_{\mathsf{PATE}} &= \mathbb{E}\left[\tau\left(X_i\right)\right] \\ \tau_{\mathsf{PATT}} &= \mathbb{E}\left[\tau\left(X_i\right) \mid D_i = 1\right] \\ \tau_{\mathsf{PATU}} &= \mathbb{E}\left[\tau\left(X_i\right) \mid D_i = 0\right]\end{aligned} \tag{8.23}$$

Multivariate regression

Define the following models for the conditional potential outcomes:

$$\begin{aligned}\mu_0(x) =& \mathbb{E}\left[Y_i^0 | X_i = x\right] \\ \mu_1(x) =& \mathbb{E}\left[Y_i^1 | X_i = x\right].\end{aligned} \tag{8.24}$$

By definition, the **average treatment effect** conditional on $X = x$ is

$$\tau(x) = \mu_1(x) - \mu_0(x). \tag{8.25}$$

Under UC, we can estimate $\mu_1(x)$ using the treated subsample and $\mu_0(x)$ using the untreated subsample to get:

$$\hat{\tau}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mu}_1(x) - \hat{\mu}_0(x)\right). \tag{8.26}$$

Multivariate regression: Linearity

In the simplest case, we assume a **linear** specification:

$$\begin{aligned} \mu_0(x) &= \alpha_0 + \beta_0' x \\ \mu_1(x) &= \alpha_1 + \beta_1' x. \end{aligned} \tag{8.27}$$

Then

$$\hat{\tau}_{\text{reg}} = \hat{\alpha}_1 - \hat{\alpha}_0 + \bar{X}'(\underbrace{\hat{\beta}_1 - \hat{\beta}_0}_{=0}), \tag{8.28}$$

which can be obtained from the coefficient of D in a regression of Y_i on $(1, D_i, X_i, D_i \cdot X_i)$.

In general, the **linear treatment effect estimator** can be expressed as

$$\hat{\tau}_{\text{reg}} = \bar{Y}_1 - \bar{Y}_0 - \left(\frac{N_0 \hat{\beta}_1}{N_0 + N_1} + \frac{N_1 \hat{\beta}_0}{N_0 + N_1} \right)' \left(\bar{X}_1 - \bar{X}_0 \right). \tag{8.29}$$

Multivariate regression: Misspecification

Thus, the average difference in **outcomes** is adjusted by the difference in **covariates**.

If averages of covariates are very different, this adjustment is **large**. In this case, results can be sensitive to minor changes in **specification**.

Unless the linear approximation is **globally** accurate, this may lead to **severe biases**.

If the average of the covariates are very different, they will be **collinear** with the treatment and exacerbate bias.

To avoid misspecification problems one can use a **matching estimator** that can estimate treatment effects more flexibly. **Regression** may be interpreted as **matching with specific weights**.

Outline

8 Matching
- Potential outcomes framework
- Identification
- Estimation
- Kernel Matching
- Inverse Probability Weighting
- Detailed Matching Application

Matching Estimators

Matching estimators impute the missing potential outcomes using outcomes of a few **neighbours** in the opposite group.

Given a **matching metric**, the researcher only has to choose the **number of matches**.

Using a single match leads to the most credible inference with the least **bias**, at the cost of lower **precision**.

This **bias-precision** trade-off is analogue to the choice of bandwidth in nonparametric regression estimation.

Matching estimators are normally used when

1. The interest is in the **ATT** rather than the **ATE**.
2. **Randomization** into treatment is not possible.
3. There is a **large pool** of potential control subjects.

Matching Estimators II

Access to a large pool of potential controls allows us to do matching **without replacement**. But then the **ordering** may matter (marginally).

Given the matched pairs, the treatment effect is estimated as the **difference in outcomes**.

Inference is based on standard methods for **differences in means**, ignoring any bias arising from the matching procedure or better **bootstrapping** methods.

Most **matching estimators** can be defined as a **weighted sum** of the outcomes according to

$$\hat{\tau} = \sum_{i=1}^{N} \lambda_i Y_i, \text{ with } \begin{cases} \sum_{i:D_i=1} \lambda_i = 1 \\ \sum_{i:D_i=0} \lambda_i = -1. \end{cases} \tag{8.30}$$

The estimators differ in the way the **weights** λ_i depend on the full vector of assignments and matrix of covariates.

Matching Estimators III

In sample $\{(Y_i, X_i, D_i)\}_{i=1}^N$, let $I_1(i)$ be the **nearest neighbour** of i. That is, $I_1(i)$ is the observation j in the dataset such that

$$\left\| X_j - X_i \right\| \leqslant \min_{k:D_k \neq D_i} \left\| X_k - X_i \right\|. \tag{8.31}$$

In general, define the m:th closest match $I_m(i)$ as the observation that satisfies

$$\sum_{l:D_l \neq D_i} \mathbb{1}\left\{ \left\| X_l - X_i \right\| \leqslant \left\| X_{l_m(i)} - X_i \right\| \right\} = m. \tag{8.32}$$

Then we can impute outcomes by

$$\hat{Y}_i(d) = \begin{cases} Y_i & \text{if } D_i = d \\ \frac{1}{M} \sum_{j \in J_M(i)} Y_j & \text{if } D_i \neq d \end{cases} \tag{8.33}$$

and estimate $\hat{\tau}_{\text{match}} = \frac{1}{N} \sum_{i=1}^N \left(\hat{Y}_i^1 - \hat{Y}_i^0 \right)$.

The Propensity Score

Problem: How to find nearest neighbour with **many** covariates! One common approach is to use the **propensity score** as a matching metric.

$$e(x) = \Pr(D_i = 1 | X_i = x). \tag{8.34}$$

Under UC (8.18), **independence** holds also after conditioning only on the propensity score (PS):

$$\begin{aligned} & D_i \perp\!\!\!\perp \left(Y_i^1, Y_i^0\right) \Big| X_i \\ & \quad \Rightarrow D_i \perp\!\!\!\perp \left(Y_i^1, Y_i^0\right) \Big| e(X_i). \end{aligned} \tag{8.35}$$

This is true since $D_i \perp\!\!\!\perp X_i | e(X_i)$. See Rosenbaum and Rubin (1983).

Thus, it is sufficient to adjust only for differences in PS between treated and control units.
Estimate and **specify** the PS e.g. using **probit**. Then choose nearest neighbour based on PS.

Caliper and Radius Matching

Oversampling: Many neighbors per treated reduces variance but increases bias (poorer matches used).

Main **weakness of NN matching**: closest neighbor(s) might be **far away**.

Caliper matching imposes **tolerance level** on maximum PS distance.

- **Bad matches** are avoided and matching quality rises.
- **Downside**: Fewer matches used $\Rightarrow$ variance increases.
- **Problem**: Difficult to know appropriate tolerance level.

Radius matching uses all comparison members within the caliper.

- The number of matches used depends on the **quality** of matches.
- Hence, it combines advantages of **oversampling** and **caliper**.

Outline

8. Matching
   - Potential outcomes framework
   - Identification
   - Estimation
   - Kernel Matching
   - Inverse Probability Weighting
   - Detailed Matching Application

Kernel Matching

The algorithms discussed so far put uniform weights on the matches.

Kernel and **local linear** estimators allow us to use weighted averages instead.

Given a kernel $K(\cdot)$, a bandwidth h, and X or the **propensity score**, the kernel estimator is

$$\begin{aligned} \widehat{\mu}_d(e) &= \sum_{i:D_i=d} Y_i \lambda_i \\ \lambda_i &= K\left(\frac{\hat{e}(X_i)-e}{h}\right) \Bigg/ \sum_{j:D_j=d} K\left(\frac{\hat{e}(X_j)-e}{h}\right) \end{aligned} \tag{8.36}$$

Local linear regression also adjust for differences in X_i which is useful to avoid **boundary point** problems.

The kernel function rarely matters, but the **bandwidth parameter** h does:

- Large bandwidth $\Rightarrow$ worse fit, lower variance.
- Small bandwidth $\Rightarrow$ better fit but higher variance.

Outline

The IPW Estimator

Recall that $\tau_{\mathsf{PATE}} = \mathbb{E}\left[Y_i^1 - Y_i^0\right] = \mathbb{E}\left[Y_i^1\right] - \mathbb{E}\left[Y_i^0\right]$. Consider

$$
\begin{aligned}
\mathbb{E}\left[Y_i^1\right] &\overset{IE}{=} \mathbb{E}\left[\mathbb{E}\left[Y_i^1 | X_i\right]\right] = \mathbb{E}\left[\frac{e\left(X_i\right)\mathbb{E}\left[Y_i^1 | X_i\right]}{e\left(X_i\right)}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}\left[D_i | X_i\right]\mathbb{E}\left[Y_i^1 | X_i\right]}{e\left(X_i\right)}\right] \overset{UC}{=} \mathbb{E}\left[\mathbb{E}\left[\left.\frac{D_i Y_i^1}{e\left(X_i\right)}\right| X_i\right]\right] \\
&\overset{IE}{=} \mathbb{E}\left[\frac{D_i Y_i^1}{e\left(X_i\right)}\right] \overset{SUTVA}{=} \mathbb{E}\left[\frac{D_i Y_i}{e\left(X_i\right)}\right]
\end{aligned}
\tag{8.37}
$$

and a similar derivation for $\mathbb{E}\left[Y_i^0\right]$. Thus,

$$
\tau_{\mathsf{PATE}} = \mathbb{E}\left[\frac{D_i Y_i}{e\left(X_i\right)} - \frac{\left(1 - D_i\right)Y_i}{1 - e\left(X_i\right)}\right]. \tag{8.38}
$$

The **inverse probability weighting estimator** is defined by

$$
\hat{\tau}_{\mathsf{IPW}} = \frac{\sum_{i=1}^N \frac{D_i Y_i}{\hat{e}(X_i)}}{\sum_{i=1}^N \frac{D_i}{\hat{e}(X_i)}} - \frac{\sum_{i=1}^N \frac{(1-D_i)Y_i}{1-\hat{e}(X_i)}}{\sum_{i=1}^N \frac{1-D_i}{1-\hat{e}(X_i)}}. \tag{8.39}
$$

Outline

8 Matching
- Potential outcomes framework
- Identification
- Estimation
- Kernel Matching
- Inverse Probability Weighting
- Detailed Matching Application

Matching Application from my own Work

Example 8.2 (Biewen, Fitzenberger, Osikominu, Paul (2014, JOLE): The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices)

- Evaluation of training schemes for the unemployed: West Germany, 2000-2003
- Effects on employment probability and earnings
- Differential effects of different programs (short-term activation vs. long-term human capital enhancement; theoretical vs. practical)
- Effect heterogeneity w.r.t. different subgroups (gender, age groups, educational groups)
- Analyze to what extend features of the data, the particular specification and the dynamic approach influence evaluation outcome: **focus here.**

Matching Application

- Data: German administrative data (IEBS)
- Programs: Short-term training (STT), Classroom further training (CFT), Practical further training (PFT)
- Account for dynamic framework (Sianesi, 2004, 2008); distinguish programs starting after elapsed unemployment: 0 to 3 months (stratum 1), 4 to 6 months (stratum 2), and 7 to 12 months (stratum 3)
- Direct comparison of different programs (Imbens, 2000; Lechner, 2001)
- Combine exact matching and propensity score matching, local linear matching.

Benchmark Specification: Matching Requirements

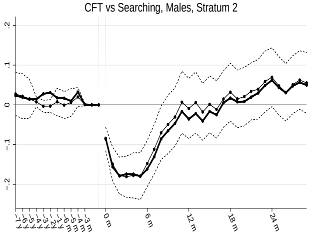
- Equality in the elapsed time in open unemployment
- Equality in the previous employment history: exactly matching on the nine sequences (1000), (1001), (1010), (1011), (1100), (1101), (1110), (1111), and (0000).
- Similarity in the pairwise propensity score (includes employment and earnings history, benefit entitlement, local labor market characteristics, rich personal characteristics etc.).
- Similarity in the calendar date of the beginning of unemployment
- Specification search for each of the 34 groups (economic considerations, significance, balancing tests). Typically 20 to 35 covariates in PS.

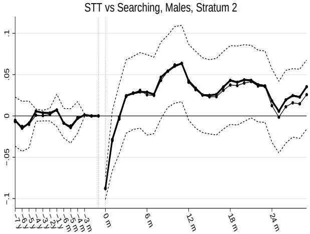
Sensitivity Analysis

- To what extent do features of our data and specification choices influence the outcome of our evaluation? What is important when choosing the econometric evaluation strategy?
- Start with benchmark specification and then sequentially drop specification features and/or information in the data.
- Aspects studied:
  - Employment History
  - Rich Personal Information and Specification Search
  - Comparison to MTG and Information on Other Programs
  - Future Participation in Other Programs

Sensitivity Analysis 1: Employment History

Step 1: No exact matching on employment histories

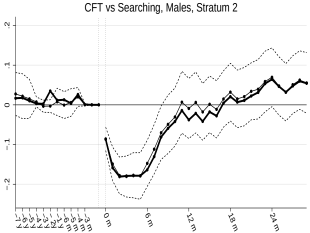


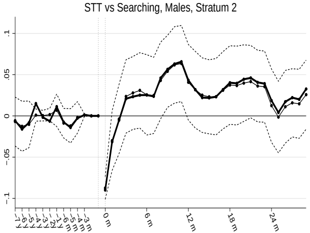


Requirement to match *exactly* on employment sequence dropped.

Sensitivity Analysis 1: Employment History

Step 2: In addition, no history sequence dummies in propensity score

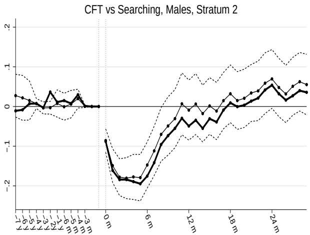


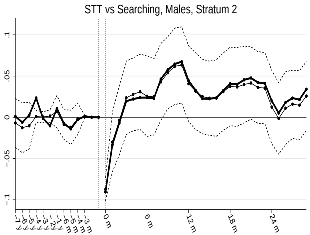


Variables with four year history sequence in addition dropped.

Sensitivity Analysis 1: Employment History

Step 3: In addition, the seven-year history information is dropped from in propensity score

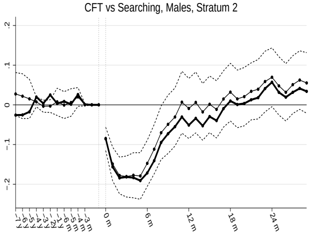


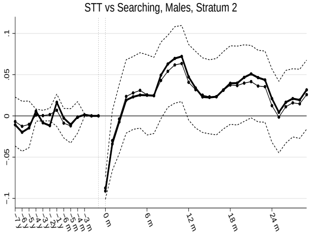


Information (on employment and earnings) going back to at most three years remains.

Sensitivity Analysis 1: Employment History

Step 4: In addition, everything related to benefit information dropped

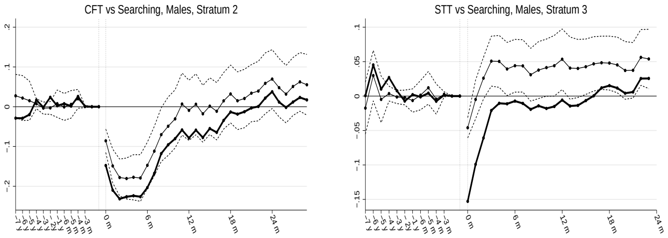




Non information on benefit claim, earlier benefit receipt etc. remains.

Sensitivity Analysis 1: Employment History

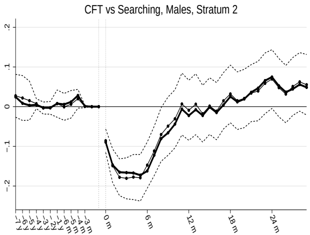
Step 5: In addition, no exact matching on unemployment duration

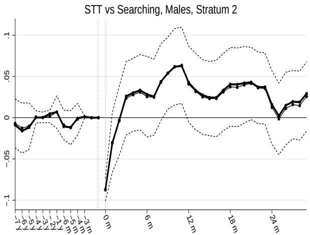


Within a stratum, those with longer and shorter elapsed unemployment durations may be matched.

Sensitivity Analysis 2: Rich Personal Information and Specification Search

Step 1: No personality variables

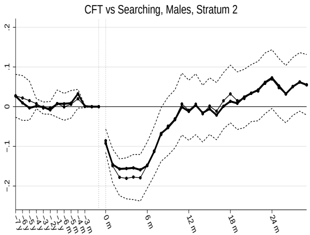


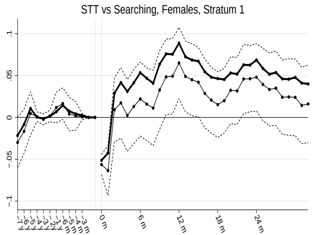


The following variables are dropped from PS: dropout or penalties in past, signs of lack of motivation in past, program with psychosocial component in past, wish to change occupation, number of job proposals.

Sensitivity Analysis 2: Rich Personal Information and Specification Search

Step 2: In addition, no rich personal information

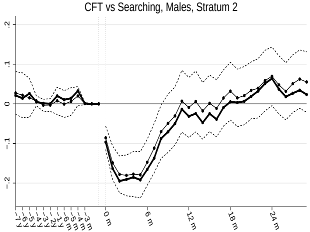


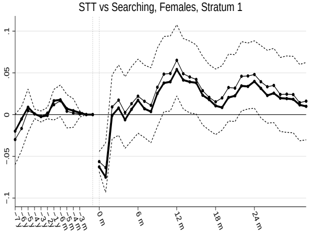


The following variables are dropped from PS: information on health, disability, number and age of dependent children, marital status, information on household type, previous part-time employment and reasons why the last job was ended.

Sensitivity Analysis 2: Rich Personal Information and Specification Search

No detailed specification search

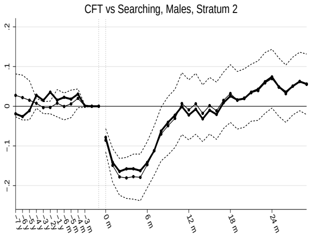


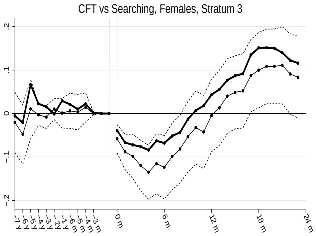


Same basic specification (same variables, no interactions etc.) for each of the 34 groups.

Sensitivity Analysis 3: Comparison with MTG and Information on Other Programs

Comparison with specification similar to MTG (Mueser et al., 2007)

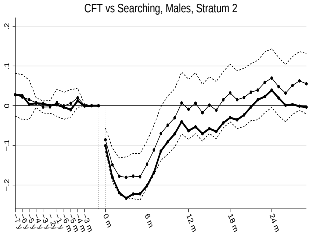


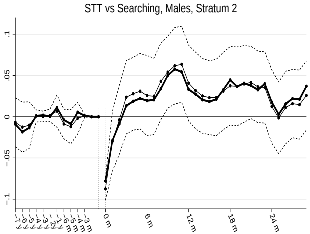


MTG: study representative for what is possible using US administrative data sources (here: Missouri); basic demographic and educational information and region, four quarters of earnings history and employment dummies based on whether earnings were positive.

Sensitivity Analysis 3: Comparison with MTG and Information on Other Programs

Ignoring information on other programs

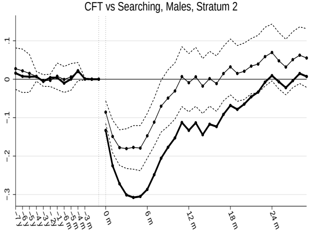


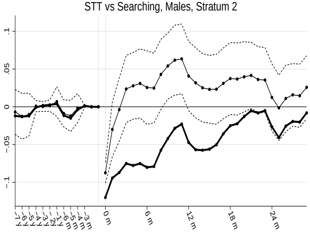


Ignores all information on programs other than the one in question; this mimics the situation in MTG who do not know if controls have participated in other programs.

Sensitivity Analysis 4: Future Participation

Excluding future participants from the beginning of stratum

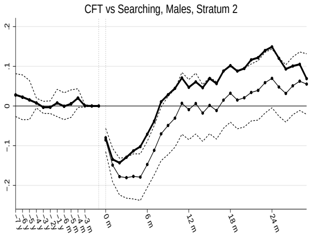


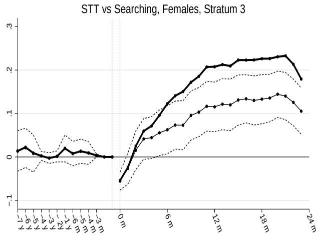


Exclude all individuals from the beginning on who will eventually (within the same unemployment spell and 35 months) participate in training or another intensive active labor market program.
Corresponds to control group design: never treated.

Sensitivity Analysis 4: Future Participation

Excluding future participants from the month they enter a program





Exclude future participants from the control group from the month they enter a program onwards. (To check whether our effects are systematically driven up because control group members are locked in future programs.)