Matching

SS 2011

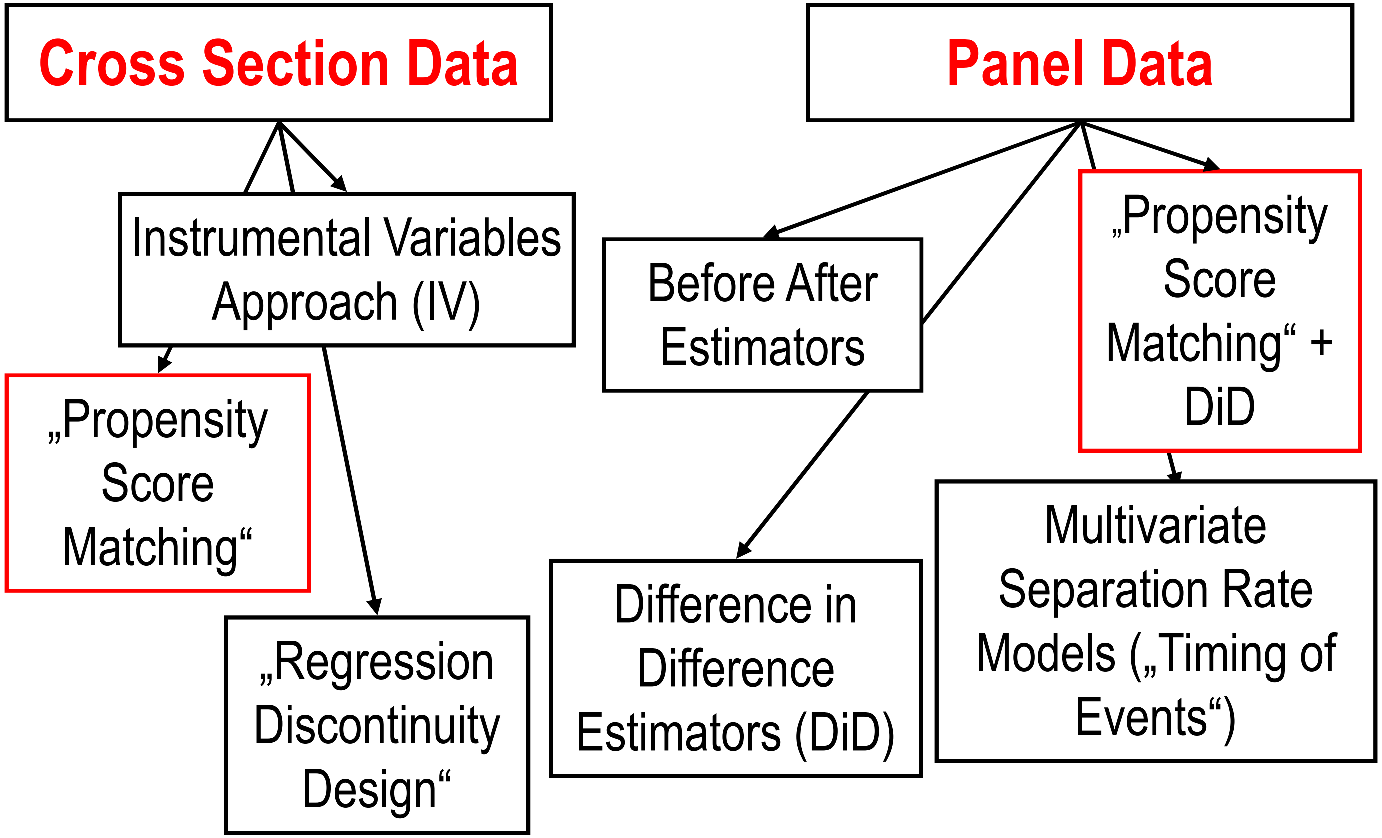
Alexander Spermann

1. Matching: Concept & Assumptions
2. Propensity Score Matching
3. Matching approaches
4. Matching vs. Regression

Main source: Caliendo (2006), Ch. 2 and Ch. 3
Supplementary reference: Angrist/Pischke (2009), Imbens/Wooldridge (2009)

1. Matching: Concept & Assumptions
Classification

Evaluation with non-experimental approaches



Cross Section Data

Instrumental Variables Approach (IV)

„Propensity Score Matching“

„Regression Discontinuity Design“

Panel Data

Before After Estimators

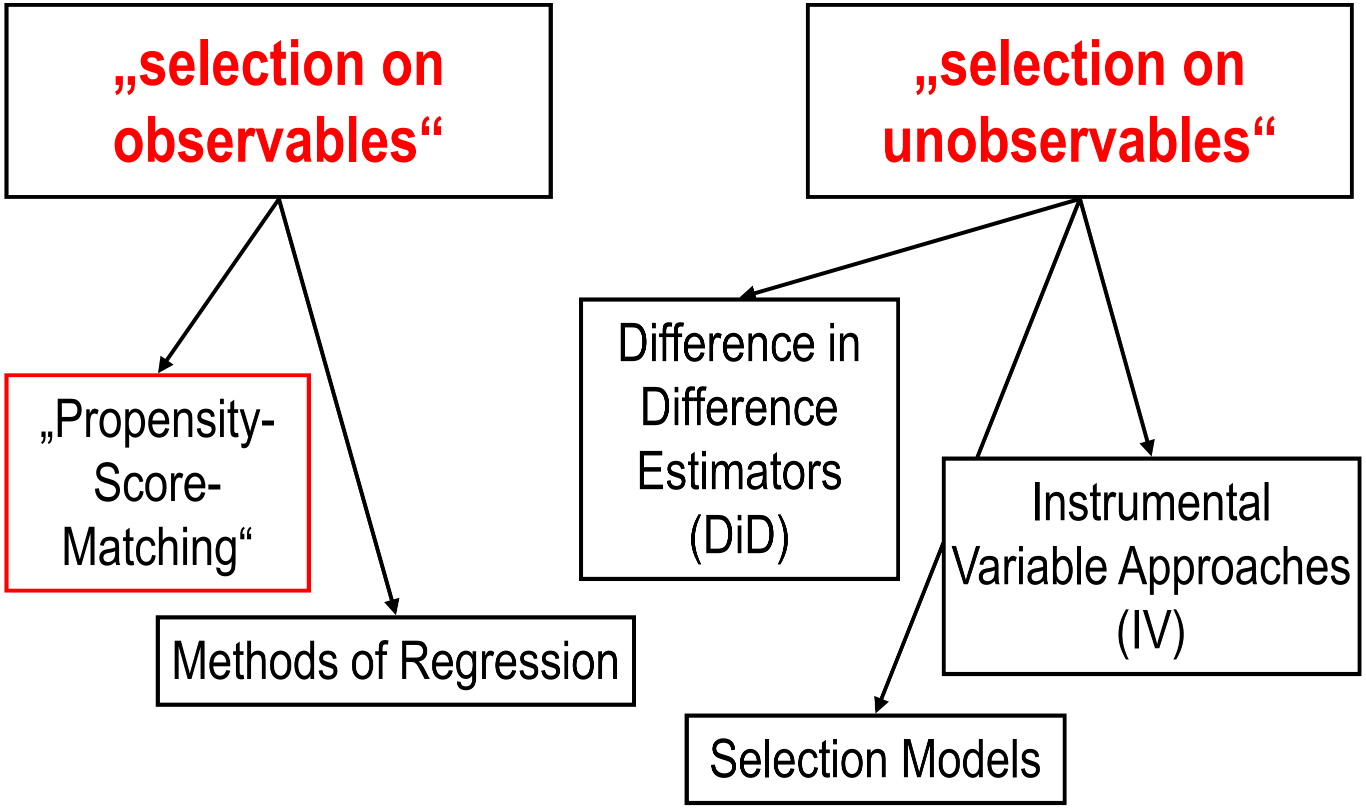
Difference in Difference Estimators (DiD)

„Propensity Score Matching“ + DiD

Multivariate Separation Rate Models („Timing of Events“)

1. Matching: Concept & Assumptions
Classification

Evaluation with non-experimental approaches



„selection on observables“

„selection on unobservables“

„Propensity-Score-Matching“

Methods of Regression

Difference in Difference Estimators (DiD)

Instrumental Variable Approaches (IV)

Selection Models

1. Matching: Concept & Assumptions

Matching

Concept:

A participant‘s outcome variable is compared to the outcome variable of one or several non participants as similar as possible to the participant.

1. Matching: Concept & Assumptions

Assumptions:

A1: Only observable factors influence participation and outcome variable simultaneously (Counter-example: motivation)
→ CMI / CIA

A2: „common support“ is given, i.e.

$$0 < P(C = 1 | x) < 1$$

1. Matching: Concept & Assumptions

Conditional Independence Assumption (CIA):

Conditional on **x**, **C** and **(y_0,y_1)** are independent.

Where **x** is the vector of all observed variables

1. Matching: Concept & Assumptions

Conditional Mean Independence Assumption (CMI):

$E[y_0 \mid x, C] = E[y_0 \mid x]$ and $E[y_1 \mid x, C] = E[y_1 \mid x]$

CIA is more restrictive than CMI. CMI is sufficient to identify ATT.

1. Matching: Concept & Assumptions

Equation

$$E[y_0 \mid x, C] = E[y_0 \mid x]$$

can also be stated as

$$E[y_0 \mid x, C=1] = E[y_0 \mid x, C=0]$$

ATT can be estimated **consistently**

→ Exact matching: for each participant find (at least) one non participant equal in x

2. Propensity Score Matching

Problem:
In case vector x is large (many variables), it is unlikely to find a similar non participant for every participant with regard to all characteristics → curse of dimensionality

Solution:

Matching of participants and non participants based on their estimated propensity scores: $p(x) \equiv P(C=1|x)$

2. Propensity Score Matching

Rosenbaum und Rubin (1983)* show that if CIA holds then:

$$E[y_0 \mid p(x), C = 1] = E[y_0 \mid p(x), C = 0]$$

Hence, matching of participants and non participants based on propensity scores is sufficient.

* Rosenbaum, P.R. und D.B. Rubin (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70 (1), S.1-55

2. Propensity Score Matching

Advantages:

- Vector x is reduced to a one-dimensional probability p(x)
- Easier to find „good matches“

Necessary Assumption for Matching:

„Common Support“

$0<p(C=1|x)<1$

→No comparison of „incomparable“ individuals

2. Propensity Score Matching

Individuals with equal characteristics of variable x should show a positive probability to be participants as well as a positive probability to be non participants.

Status of participation cannot be predicted using x.

2. Propensity Score Matching Procedure

Procedure

- Estimation of a model for participation C conditioned on explanatory vector x (e.g.: Probit)
- Prediction of individual propensity scores based on the estimation
- Estimation of ATT as:

$$A\hat{T}T = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(y_{1i} - \sum_{j=1}^{N_0} w_{ij} y_{0j} \right)$$

2. Propensity Score Matching Procedure

A weighted average of all non participants‘ outcome variables is subtracted from every participants outcome variable.
Where:

N_1	Number of participants
N_0	Number of non participants
i	Index of participants
j	Index of non participants
w_{ij}	Weights

$$where\, w_{ij} \in [0,1]\; and \sum_{j=1}^{N_0} w_{ij} = 1$$

2. Propensity Score Matching
Procedure

- The smaller the difference in estimated propensity scores $|p_i-p_j|$ or equivalently the more similar participant i and non participant j are, the higher the weights

- Weights of different Matching approaches differ:

➢ „Nearest-Neighbour-Matching“
➢ „Caliper Matching“
➢ „Kernel-based Matching“

3. Matching – Approaches

„Nearest-Neighbour-Matching“

- There is only one member of the control group per participant: The non participant minimizing $|p_i-p_j|$

- This „next neighbour‘s“ weight is set to 1

3. Matching – Approaches

„Caliper Matching“

- Participant i is only included in the ATT calculation in case a next neighbour within a priori specified range $|p_i-p_j|$ can be found

→ Modification of „Nearest-Neighbour-Matching“

→ Reduces the probability of „bad matches“

3. Matching – Approaches

„Kernel based Matching“

- Use of several or all non participants as control group for every participant

- Weight for non participant j is a negative function of $|p_i - p_j|$.

3. Matching – Approaches

→<u>All available information is included</u>

→Reduction of the estimation‘s variance

→Danger of higher selectivity bias as a participant can be assigned to unsimilar non participants

4. Matching vs. Regression

Note: A linear regression model following the equation

$$y_i = x_i^{'}\beta + \beta_c C + u_i$$

is also conditioned on observable variables which do have an influence on participation and the outcome variable.

4. Matching vs. Regression

Disadvantages compared to Matching:

- No functional form assumptioon
- Common support unclear
- Heterogeneous treatment effect cannot be measured