Simplifying Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

University of Wisconsin, 2/7/2021

\(^1\text{GaryKing.org}\)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:
   ⇝ "Why Propensity Scores Should Not Be Used for Matching" (PA, 2019. Gary King, Richard Nielsen)

2. Do powerful methods have to be complicated?
   ⇝ "Causal Inference Without Balance Checking: Coarsened Exact Matching" (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

3. Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously:
   ⇝ "The Balance-Sample Size Frontier in Matching Methods for Causal Inference" (AJPS; Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:


2. Do powerful methods have to be complicated?

[ -> “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

3. Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously:

[ -> “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (AJPS; Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:

2. Do powerful methods have to be complicated?
   ⇝ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

3. Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously:
   ⇝ “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (AJPS; Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:
2. Do powerful methods have to be complicated?
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:
   \[ \Rightarrow \] “Why Propensity Scores Should Not Be Used for Matching” \((PA, 2019. \ Gary \ King, \ Richard \ Nielsen)\)

2. Do powerful methods have to be complicated?
   \[ \Rightarrow \] “Causal Inference Without Balance Checking: Coarsened Exact Matching” \((PA, 2011. \ Stefano \ Iacus, \ Gary \ King, \ and \ Giuseppe \ Porro)\)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:

2. Do powerful methods have to be complicated?
   ~ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

3. Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously:
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 336,000 articles!) sounds magical:

2. Do powerful methods have to be complicated?
   ⇝ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

3. Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously:
   ⇝ “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (AJPS; Gary King, Christopher Lucas and Richard Nielsen)
Matching to Reduce Model Dependence
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing the relationship between education and outcome](image-url)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
The Problems Matching Solves

- Qualitative choice from unbiased estimates = biased estimator
- e.g., Choosing from results of 50 randomized experiments
- Choosing based on "plausibility" is probably worse
- conscientious effort doesn't avoid biases (Banaji 2013)
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)
- "Teaching psychology is mostly a waste of time" (Kahneman 2011)
The Problems Matching Solves

Without Matching:

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
- Choosing based on "plausibility" is probably worse
- Conscientious effort doesn’t avoid biases (Banaji 2013)
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)
- “Teaching psychology is mostly a waste of time” (Kahneman 2011)
The Problems Matching Solves

Without Matching:

Imbalance
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence
The Problems Matching Solves

Without Matching:

Imbalance $\rightarrow$ Model Dependence $\rightarrow$ Researcher discretion

- Qualitative choice from unbiased estimates = biased estimator
- e.g., Choosing from results of 50 randomized experiments
- Choosing based on “plausibility” is probably worse
- Conscientious effort doesn’t avoid biases (Banaji 2013)
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)
- “Teaching psychology is mostly a waste of time” (Kahneman 2011)
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse

- conscientious effort doesn’t avoid biases (Banaji 2013)
The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse \cite{effrt}

- conscientious effort doesn’t avoid biases \cite{acc}

- People do not have easy access to their own mental processes or feedback to avoid the problem \cite{exprt}

\cite{Banaji2013}
\cite{WilsonBrekke1994}
\cite{Tetlock2005}
\cite{Kahneman2011}
The Problems Matching Solves

Without Matching:

Imbalance $\rightarrow$ Model Dependence $\rightarrow$ Researcher discretion $\rightarrow$ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse

- Conscientious effort doesn’t avoid biases (Banaji 2013)

- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)

- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)
The Problems Matching Solves

Without Matching:

Imbalance \(\sim\) Model Dependence \(\sim\) Researcher discretion \(\sim\) Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse

- Conscientious effort doesn’t avoid biases (Banaji 2013)

- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)

- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)

- “Teaching psychology is mostly a waste of time” (Kahneman 2011)
The Problems Matching Solves

Without Matching:

Imbalance ⇝ Model Dependence ⇝ Researcher discretion ⇝ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

• Qualitative choice from unbiased estimates = biased estimator
  • e.g., Choosing from results of 50 randomized experiments
  • Choosing based on "plausibility" is probably worse

• Conscientious effort doesn't avoid biases (Banaji 2013)

• People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)

• Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)

• "Teaching psychology is mostly a waste of time" (Kahneman 2011)
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\rightarrow$ Model Dependence $\rightarrow$ Researcher discretion $\rightarrow$ Bias

A central project of statistics: Automating away human discretion
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders

Treatment Effect for treated observation $i$:

$$\text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$

- Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control

Quantities of Interest:

1. SATT: Sample Average Treatment effect on the Treated:
   $$\text{SATT} = \text{Mean}_{i \in \{T_i = 1\}}(\text{TE}_i)$$
2. FSATT: Feasible SATT (prune badly matched treateds too)

- Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

$$TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$

- Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control

Quantity of Interest:

1. SATT: Sample Average Treatment effect on the Treated:
   $$SATT = \text{Mean}_{i \in \{T_i = 1\}} (TE_i)$$

2. FSATT: Feasible SATT (prune badly matched treateds too)

Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching

Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
What’s Matching?

- \( Y_i \) dep var, \( T_i \) (1=treated, 0=control), \( X_i \) confounders
- Treatment Effect for treated observation \( i \):
  \[
  \text{TE}_i = Y_i(1) - Y_i(0)
  \]
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

$$TE_i = Y_i(1) - Y_i(0)$$

= observed $-$ unobserved
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

$$TE_i = Y_i - Y_i(0)$$

$= \text{observed} - \text{unobserved}$
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

\[ TE_i = Y_i - Y_i(0) \]

= observed $-$ unobserved

- Estimate $Y_i(0)$ with $Y_j$ from a matched $(X_i \approx X_j)$ control
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  \text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]
- Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated: $\text{SATT} = \text{Mean}_{i \in \{T_i = 1\}} (\text{TE}_i)$
  2. FSATT: Feasible SATT (prune badly matched treateds too)

Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching

Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

$$\text{TE}_i = Y_i - Y_i(0)$$

$= \text{observed} - \text{unobserved}$

- Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \text{Mean} \left( \text{TE}_i \right)_{i \in \{T_i=1\}}$$
What’s Matching?

• $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
• Treatment Effect for treated observation $i$:

$$TE_i = Y_i - Y_i(0)$$

$= \text{observed} - \text{unobserved}$

• Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control
• Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \text{Mean}_{i \in \{T_i=1\}} (TE_i)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)
What’s Matching?

• $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
• Treatment Effect for treated observation $i$:

$$TE_i = Y_i - Y_i(0)$$

$= \text{observed} - \text{unobserved}$

• Estimate $Y_i(0)$ with $Y_j$ from a matched ($X_i \approx X_j$) control
• Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{Mean}_{\{T_i=1\}} (TE_i)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)
• Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching
What’s Matching?

- \( Y_i \) dep var, \( T_i \) (1=treated, 0=control), \( X_i \) confounders
- Treatment Effect for treated observation \( i \):
  \[
  \text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]
- Estimate \( Y_i(0) \) with \( Y_j \) from a matched (\( X_i \approx X_j \)) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{Mean } (\text{TE}_i)_{i \in \{T_i=1\}}
     \]
  2. FSATT: Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
Matching: Finding Hidden Randomized Experiments

Types of Experiments

- **Balance**
  - **Covariates:**
    - **Complete Randomization**
    - **Fully Blocked**
  - **Observed**
    - **On average**
  - **Unobserved**
    - **On average**

→ Fully blocked dominates complete randomization for:
  - Imbalance,
  - Model dependence,
  - Power,
  - Efficiency,
  - Bias,
  - Research costs,
  - Robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

Covariates:
- Complete Randomization
- Fully Blocked

Observed
- On average
- Imbalance
- Model dependence
- Power
- Efficiency
- Bias
- Research costs
- Robustness

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

- Complete Randomization
- Fully Blocked
- Observed

⇝ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.

E.g., Imai, King, Nall 2009: SEs 60% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

Complete Randomization

⇝

Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

• PSM: complete randomization
• Other methods: fully blocked
• Other matching methods dominate PSM
## Matching: Finding Hidden Randomized Experiments

### Types of Experiments

<table>
<thead>
<tr>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
</table>

On average, `fully blocked` dominates `complete randomization` for:
- imbalance,
- model dependence,
- power,
- efficiency,
- bias,
- research costs,
- robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance
Covariates:
Observed
Unobserved

Complete Randomization
Fully Blocked

⇝

On average

Full blocked dominates complete randomization for:
imbalance,
model dependence,
power,
efficiency,
bias,
research costs,
robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>Observed</td>
<td>On average</td>
</tr>
</tbody>
</table>

E.g., Imai, King, Nall 2009: SEs 60% smaller!
Matching: Finding Hidden Randomized Experiments

### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

E.g., Imai, King, Nall 2009: SEs 600% smaller!
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\(\Rightarrow\) Fully blocked dominates complete randomization

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \Rightarrow \text{Fully blocked dominates complete randomization for:} \]
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\Rightarrow \text{Fully blocked dominates complete randomization for: imbalance,}\]
### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covariates:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[ \Rightarrow \text{Fully blocked dominates complete randomization for:} \]

imbalance, model dependence,
# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

⇴ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>Unobserved</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[ \Rightarrow \text{Fully blocked dominates complete randomization for:} \]
imbalance, model dependence, power, efficiency,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\Rightarrow\] Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\rightarrow\) Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

$\rightsquigarrow$ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\(\leadsto\) Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td>On average</td>
</tr>
</tbody>
</table>

\(\Rightarrow\) Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

~ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\(~\) Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
- Other methods: fully blocked
Matching: Finding Hidden Randomized Experiments

**Types of Experiments**

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td>On average</td>
</tr>
</tbody>
</table>

$\rightarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

**Goal of Each Matching Method (in Observational Data)**

- PSM: *complete randomization*
- Other methods: *fully blocked*
- Other matching methods dominate PSM
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th></th>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
- Other methods: fully blocked
- Other matching methods dominate PSM (wait, it gets worse)
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - \( \text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)} \)
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if \( \text{Distance} > \text{caliper} \)
   - (Many adjustments available to this basic method)

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   \[ \text{Distance}(\mathbf{X}_c, \mathbf{X}_t) = \sqrt{(\mathbf{X}_c - \mathbf{X}_t)' S^{-1} (\mathbf{X}_c - \mathbf{X}_t)} \]
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper
   - (Many adjustments available to this basic method)

2. Estimation
   Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance\((X_c, X_t) = \sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}\)

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance\((X_c, X_t) = \sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}\)
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching  
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)  
   - Distance\((X_c, X_t) = \sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}\)  
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)  
   - Match each treated unit to the nearest control unit

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Distance($X_c, X_t$) = $\sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching  
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Distance($X_c, X_t$) = $\sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $> caliper$

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Distance\( (X_c, X_t) = \sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)} \)
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $>$ *caliper*
   - (Many adjustments available to this basic method)

2. **Estimation** Difference in means or a model
Mahalanobis Distance Matching

Age

Education (years)
Mahalanobis Distance Matching
Mahalanobis Distance Matching

Education (years)

<table>
<thead>
<tr>
<th>Age</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mahalanobis Distance Matching

Age
12 14 16 18 20 22 24 26 28

Education (years)
20 30 40 50 60 70 80

T
T T
TT TT T TT
TTT TT
T
T TT
C
C
C C
CC
C
C
C CCC
C
C
CCCC C CCC C
CCCC
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$,
     $C(X)$
   - Sort observations into strata, each with unique values of $C(X)$
   - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. Estimation
   - Difference in means or a model
   - Weight controls in each stratum to equal treateds

10 / 26
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Temporarily coarsen $X$ as much as you’re willing

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching  
(Most powerful easy-to-use approach)  
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)  
   • Temporarily coarsen $X$ as much as you’re willing  
     • e.g., Education (grade school, high school, college, graduate)  
   • Apply exact matching to the coarsened $X$, $C(X)$  
     • Sort observations into strata, each with unique values of $C(X)$  
     • Prune any stratum with 0 treated or 0 control units

2. Estimation  
   Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. **Estimation** Difference in means or a model
   - Weight controls in each stratum to equal treateds
Coarsened Exact Matching
Coarsened Exact Matching

Education

HS  BA  MA  PhD  2nd PhD

Don't trust anyone over 30
The Big 40
Senior Discounts
Retirement
Old

CCC C
CC CC
C CC C CCC CCCC CCC CC CCC CCCCCC
C CCC CC C
T T
T T
TT TT T TT
TTT TT
T
T TT
Coarsened Exact Matching

Education
HS BA MA PhD 2nd PhD

Drinking age

Don't trust anyone over 30
The Big 40
Senior Discounts
Retirement
Old
Coarsened Exact Matching

Education

HS BA MA PhD 2nd PhD

Drinking age

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

Education

HS BA MA PhD 2nd PhD

Senior Discounts

The Big 40

Retirement

Old
Coarsened Exact Matching

Education vs. Age

Age
12 14 16 18 20 22 24 26 28

Education
20 30 40 50 60 70 80
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching

![Graph showing data points with ages ranging from 12 to 28 and education levels ranging from 20 to 80.](image-url)
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching

![Scatter plot showing the relationship between age and education.](scatter_plot.png)
Method 3: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar $\pi_i = \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}$
   - Distance($X_c$, $X_t$) = $|\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper
   - (Many adjustments available to this basic method)

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1 + e^{-X_i\beta}} \]

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}} \]
   - Distance($X_c$, $X_t$) = $|\pi_c - \pi_t|$

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1 + e^{-X_i\beta}} \]
   - Distance($X_c, X_t$) = $|\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess** (Matching)
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}} \]
   - Distance($X_c, X_t$) = $|\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1+e^{-x_i\beta}} \]
   - Distance($X_c, X_t$) = $|\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $>$ caliper

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1 + e^{-X_i \beta}} \]
   - Distance($X_c$, $X_t$) = $|\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $> caliper$
   - (Many adjustments available to this basic method)

2. Estimation Difference in means or a model
Propensity Score Matching

Age vs. Education (years)

- C: Control
- T: Treatment
Propensity Score Matching

Age

20 30 40 50 60 70 80

C C CC C C C C C CC C CC C CC C C C C C CC C C C C T T T T T T T T T T T T

Education (years)

12 16 20 24 28

Propensity Score
Propensity Score Matching

Education (years) vs. Age

Propensity Score
Propensity Score Matching

Age

Education (years)

Propensity Score

1
0

12 16 20 24 28
20
30
40
50
60
70
80
1
0
Propensity Score Matching

Age
12 16 20 24 ...

Education (years)

Propensity Score

Education (years)

Propensity Score

0
1
Propensity Score Matching

The graph shows the relationship between age and education (years), with propensity scores plotted on the vertical axis. The data points are connected by lines, indicating matching pairs. Each point represents an individual with their age on the horizontal axis and education years on the vertical axis. The propensity scores are indicated by the letter 'C' or 'T', with 'C' typically representing the control group and 'T' the treatment group.
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

Age

Education (years)
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

Education (years)

Age

12 16 20 24 28

20
30
40
50
60
70
80

C
C
C C
C
C
C
C
C ...
T
T
T
T
T
T
T
T
T
T
T
T
T
T
T

TT
T
T
T
TT
TT

Education (years)
Best Case: Propensity Score Matching is Suboptimal
Random Pruning Increases Imbalance

- **Random pruning**: pruning process is independent of \(X\).

- **Discrete example**
  - **Sex-balanced dataset**: treated \(M_t\), \(F_t\), controls \(M_c\), \(F_c\).
  - Randomly prune 1 treated & 1 control \(\Rightarrow \) 4 possible datasets: 2 balanced \{\(M_t, M_c\)\}, \{\(F_t, F_c\)\}; 2 imbalanced \{\(M_t, F_c\)\}, \{\(F_t, M_c\)\} \(\Rightarrow \) random pruning increases imbalance.

- **Continuous example**
  - Dataset: \(T \in \{0, 1\}\) randomly assigned; \(X\) any fixed variable; with \(n\) units.
  - Imbalance measure, squared difference in means: \(d^2 = (\bar{X}_t - \bar{X}_c)^2\).
  - \(E(d^2) = V(d) \propto \frac{1}{n}\) (because \(E(d) = 0\)).
  - Random pruning \(\Rightarrow n\) declines \(\Rightarrow E(d^2) increases\).

- Result is completely general (see math in the paper).
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

• "Random pruning": pruning process is independent of \( X \)

• Discrete example

  • Sex-balanced dataset: treateds \( M_t, F_t \), controls \( M_c, F_c \)

  • Randomly prune 1 treated & 1 control

    \( \rightarrow \) 4 possible datasets:

    2 balanced \( \{ M_t, M_c \}, \{ F_t, F_c \} \)

    2 imbalanced \( \{ M_t, F_c \}, \{ F_t, M_c \} \)

  \( \Rightarrow \) random pruning increases imbalance

• Continuous example

  • Dataset: \( T \in \{ 0, 1 \} \) randomly assigned; \( X \) any fixed variable; with \( n \) units

  • Imbalance measure, squared difference in means:
    
    \[ d^2 = (\bar{X}_t - \bar{X}_c)^2 \]

  \( \Rightarrow \)

  \[ E(d^2) = V(d) \propto 1/n \] (because \( E(d) = 0 \))

  • Random pruning \( \rightarrow n \) declines \( \rightarrow E(d^2) \) increases

  \( \Rightarrow \) random pruning increases imbalance

• Result is completely general (see math in the paper)
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$

Discrete example

- Sex-balanced dataset: treated $M_t$, $F_t$, controls $M_c$, $F_c$
- Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets:
  - 2 balanced: $\{M_t, M_c\}$, $\{F_t, F_c\}$
  - 2 imbalanced: $\{M_t, F_c\}$, $\{F_t, M_c\}$

- $\Rightarrow$ random pruning increases imbalance

Continuous example

- Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
- Imbalance measure, squared difference in means:
  $$d^2 = (\bar{X}_t - \bar{X}_c)^2$$
- $E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)

- Random pruning $\Rightarrow n$ declines $\Rightarrow E(d^2)$ increases

- $\Rightarrow$ random pruning increases imbalance

- Result is completely general (see math in the paper)
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example

Sex-balanced dataset: treated $M_t$, $F_t$, controls $M_c$, $F_c$

Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets: 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$, 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$

$\Rightarrow$ random pruning increases imbalance

Continuous example

Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units

Imbalance measure, squared difference in means: $d^2 = (\bar{X}_t - \bar{X}_c)^2$

$E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)

$\Rightarrow$ random pruning $\Rightarrow n$ declines $\Rightarrow E(d^2)$ increases

$\Rightarrow$ random pruning increases imbalance

Result is completely general (see math in the paper)
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example
  - Sex-balanced dataset: treated $M_t$, $F_t$, controls $M_c$, $F_c$

Random pruning $\Rightarrow$ $n$ declines $\Rightarrow$ $E(\bar{X}_t - \bar{X}_c)^2$ increases

Result is completely general (see math in the paper)
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$

• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of X

• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\sim$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  • $\implies$ random pruning increases imbalance
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$
• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  • $\implies$ random pruning increases imbalance
• Continuous example
  • Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  • Imbalance measure, squared difference in means:
    $d^2 = (\bar{X}_t - \bar{X}_c)^2$
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\rightarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Imbalance measure, squared difference in means:
    $d^2 = (\bar{X}_t - \bar{X}_c)^2$
  - $E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
  - Dataset: $T \in \{0,1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Imbalance measure, squared difference in means:
    $d^2 = (\bar{X}_t - \bar{X}_c)^2$
    $E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)
  - Random pruning $\leadsto n$ declines $\leadsto E(d^2)$ increases
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$

• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  • $\implies$ random pruning increases imbalance

• Continuous example
  • Dataset: $T \in \{0,1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  • Imbalance measure, squared difference in means:
    $d^2 = (\bar{X}_t - \bar{X}_c)^2$
  • $E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)
  • Random pruning $\rightsquigarrow$ $n$ declines $\rightsquigarrow E(d^2)$ increases
  • $\implies$ random pruning increases imbalance

Result is completely general (see math in the paper)
Random Pruning Increases Imbalance
 Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t$, $F_t$, controls $M_c$, $F_c$
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Imbalance measure, squared difference in means:
    $d^2 = (\bar{X}_t - \bar{X}_c)^2$
  - $E(d^2) = V(d) \propto 1/n$ (because $E(d) = 0$)
  - Random pruning $\rightsquigarrow n$ declines $\rightsquigarrow E(d^2)$ increases
  - $\implies$ random pruning increases imbalance
- Result is completely general (see math in the paper)
PSM’s Statistical Properties

1. Low Standards:
   - Sometimes helps, never optimizes
   - Efficient relative to complete randomization, but inefficient relative to (the more powerful) full blocking
   - Other methods usually dominate:

2. The PSM Paradox:
   - When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning)
     \[ \hat{\pi}_c \approx 0 \] (or constant within strata)
     \[ \Rightarrow \text{pruning at random} \Rightarrow \text{Imbalance} \Rightarrow \text{Inefficency} \Rightarrow \text{Model dependence} \Rightarrow \text{Bias} \]
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem?
     Nope.
     The PSM Paradox gets worse with more covariates.
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking

\[
\pi_c = \pi_t \Rightarrow \pi_c = \pi_t \neq \pi_t^* \Rightarrow X_c = X_t
\]
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:

PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \]
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \quad \text{but} \quad \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning)
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata)
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) pruning at random \( \rightsquigarrow \) Imbalance
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) Imbalance \( \sim \) Inefficency
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \not\leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \not\leadsto \) pruning at random \( \not\leadsto \) Imbalance \( \not\leadsto \) Inefficiency \( \not\leadsto \) Model dependence
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficiency \( \leadsto \) Model dependence \( \leadsto \) Bias
1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
   \[
   X_c = X_t \implies \pi_c = \pi_t \quad \text{but} \\
   \pi_c = \pi_t \not\implies X_c = X_t
   \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \(\implies\) all \(\hat{\pi} \approx 0.5\) (or constant within strata) \(\implies\) pruning at random \(\implies\) Imbalance \(\implies\) Inefficiency \(\implies\) Model dependence \(\implies\) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
PSM’s Statistical Properties

1. **Low Standards: Sometimes helps, never optimizes**
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox: When you do “better,” you do worse**
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficency \( \leadsto \) Model dependence \( \leadsto \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem?
1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) pruning at random \( \rightsquigarrow \) Imbalance \( \rightsquigarrow \) Inefficency \( \rightsquigarrow \) Model dependence \( \rightsquigarrow \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope.
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficency \( \leadsto \) Model dependence \( \leadsto \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates
PSM is Blind Where Other Methods Can See
PSM is Blind Where Other Methods Can See
PSM is Blind Where Other Methods Can See
What Does PSM Match?

MDM Matches

PSM Matches

Controls: \( X_1, X_2 \sim \text{Uniform}(0,5) \)
Treated: \( X_1, X_2 \sim \text{Uniform}(1,6) \)
PSM Increases Model Dependence & Bias

Model Dependence

Bias

\[ Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i \]
\[ \epsilon_i \sim N(0, 1) \]
The Propensity Score Paradox in Real Data
Finkel et al. (JOP, 2012)

Nielsen et al. (AJPS, 2011)

The Propensity Score Paradox in Real Data
The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012)

Nielsen et al. (AJPS, 2011)

Similar pattern for > 20 other real data sets we checked
The Matching Frontier

- Frontier = matched dataset with lowest imbalance for each
- Bias-Variance trade off $\Rightarrow$ Imbalance-n Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible; you see everything optimal
- Choose an imbalance metric, then run.
The Matching Frontier

- Frontier = matched dataset with lowest imbalance for each $n$
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
- Simple to use
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
- Simple to use
- No need to choose or use a matching method
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
The Matching Frontier

- Frontier = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each \( n \)
- Bias-Variance trade off \( \leadsto \) Imbalance-\( n \) Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible; you see everything optimal
The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible; you see everything optimal
- Choose an imbalance metric, then run.
How hard is the frontier to calculate?

Consider 1 point on the SATT frontier:

- Start with matrix of $N$ control units $X_0$
- Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
- Choose subset with lowest imbalance

Evaluations needed to compute the entire frontier:

- $\binom{N}{n}$ evaluations for each sample size $n = N, \ldots, 1$
- The combination is the (gargantuan) “power set”
- e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

$\Rightarrow$ It’s hard to calculate!

We develop algorithms for the (optimal) frontier which:

- runs very fast
- operate as “greedy” but we prove are optimal
- do not require evaluating every subset
- work with very large data sets
- is the exact frontier (no approximation or estimation)

$\Rightarrow$ It’s easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:

  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance
  - Evaluations needed to compute the entire frontier:
    - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s hard to calculate!

  - We develop algorithms for the (optimal) frontier which:
    - runs very fast
    - operate as “greedy” but we prove are optimal
    - do not require evaluating every subset
    - work with very large data sets
    - is the exact frontier (no approximation or estimation)
  - $\Rightarrow$ It’s easy to calculate!
How hard is the frontier to calculate?

• Consider 1 point on the SATT frontier:
  • Start with matrix of $N$ control units $X_0$

• Evaluations needed to compute the entire frontier:
  • $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$

• The combination is the (gargantuan) “power set”

• e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

• $\Rightarrow$ It’s hard to calculate!

• We develop algorithms for the (optimal) frontier which:
  • runs very fast
  • operate as “greedy” but we prove are optimal
  • do not require evaluating every subset
  • work with very large data sets
  • is the exact frontier (no approximation or estimation)

• $\Rightarrow$ It’s easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

- $\Rightarrow$ It's hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - run very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)

- $\Rightarrow$ It's easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

- It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - Runs very fast
  - Operate as “greedy” but we prove are optimal
  - Do not require evaluating every subset
  - Work with very large data sets

- It’s easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$

$\Rightarrow$ It's hard to calculate!

We develop algorithms for the (optimal) frontier which:
- runs very fast
- operate as "greedy" but we prove are optimal
- do not require evaluating every subset
- work with very large data sets
- is the exact frontier (no approximation or estimation)

$\Rightarrow$ It's easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of \( N \) control units \( X_0 \)
  - Calculate imbalance for all \( \binom{N}{n} \) subsets of rows of \( X_0 \)
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - \( \binom{N}{n} \) evaluations for each sample size \( n = N, N - 1, \ldots, 1 \)
  - The combination is the (gargantuan) “power set”
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

$\Rightarrow$ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)

$\Rightarrow$ It’s easy to calculate!
How hard is the frontier to calculate?

• Consider 1 point on the SATT frontier:
  • Start with matrix of $N$ control units $X_0$
  • Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  • Choose subset with lowest imbalance

• Evaluations needed to compute the entire frontier:
  • $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  • The combination is the (gargantuan) “power set”
  • e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  • $\Rightarrow$ It’s hard to calculate!

We develop algorithms for the (optimal) frontier which:
  • runs very fast
  • operate as “greedy” but we prove are optimal
  • do not require evaluating every subset
  • work with very large data sets
  • is the exact frontier (no approximation or estimation)

$\Rightarrow$ It’s easy to calculate!
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of \(N\) control units \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - \(\binom{N}{n}\) evaluations for each sample size \(n = N, N - 1, \ldots, 1\)
  - The combination is the (gargantuan) “power set”
  - e.g., \(N > 300\) requires more imbalance evaluations than elementary particles in the universe
  - \(\Rightarrow\) It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all ${N \choose n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - ${N \choose n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s **hard** to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - ⇾ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of $N$ control units $X_0$
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\Rightarrow$ It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)
  - $\Rightarrow$ It’s easy to calculate!
Job Training Data: Frontier and Causal Estimates

- 185 Ts; pruning most 16,252 Cs won’t increase variance much
- Huge bias-variance trade-off after pruning most Cs
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed
Conclusions

Propensity score matching:
- Approximates complete, not fully blocked, experiments
- Ignores information; exacerbates model dependence
- Some mistakes with PSM:
  - Controlling for irrelevant covariates;
  - Adjusting experimental data;
  - Reestimating propensity score
    after eliminating noncommon support;
  - 1/4 caliper on propensity score;
  - Not switching to other methods.

A Simple and Powerful Method: CEM
A New General Approach: The Matching Frontier
- Fast; easy; no iteration; Software: MatchingFrontier
- No need to choose among matching methods
- Optimal results from your choice of imbalance metric

⇝ Using more information is simpler and more powerful
Conclusions

- Propensity score matching:

  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; adjusting experimental data; reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; not switching to other methods.

- A Simple and Powerful Method: CEM

- A New General Approach: The Matching Frontier

  - Fast; easy; no iteration; Software: MatchingFrontier

  - No need to choose among matching methods

  - Optimal results from your choice of imbalance metric

  - Using more information is simpler and more powerful
Conclusions

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM:
Conclusions

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates;
Conclusions

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data;
Conclusions

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support;
Conclusions

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score;
Conclusions

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- A Simple and Powerful Method: CEM
Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier
Conclusions

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- **A Simple and Powerful Method:** CEM

- **A New General Approach:** The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier

25 / 26
Conclusions

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A Simple and Powerful Method: CEM

• A New General Approach: The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
Conclusions

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- **A Simple and Powerful Method:** CEM

- **A New General Approach:** The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
Conclusions

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A Simple and Powerful Method: CEM

• A New General Approach: The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Using more information is simpler and more powerful
For more information, articles, & software

GaryKing.org