Simplifying Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

University of Pennsylvania, APPC, 4/1/2016
1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical: “Why Propensity Scores Should Not Be Used for Matching” (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” (In press, AJPS; Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:

   “Why Propensity Scores Should Not Be Used for Matching” (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” (In press, AJPS; Gary King, Christopher Lucas and Richard Nielsen)
1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
   \[\rightsquigarrow\] “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
    ~ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

2. Do powerful methods have to be complicated?
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
   ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)
2. Do powerful methods have to be complicated?
   ⇝ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano M Iacus, Gary King, and Giuseppe Porro)
3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
   ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

2. Do powerful methods have to be complicated?
   ⇝ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano M 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 53,200 articles!) sounds magical:
   ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

2. Do powerful methods have to be complicated?
   ⇝ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano M 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” (In press, 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*)

![Graph showing data points scattered across the Education (years) and Outcome axes. The x-axis ranges from 12 to 28 years, and the y-axis ranges from 0 to 12. The data points are marked with T symbols.]
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*)

![Graph showing the relationship between education (years) and outcome. The graph illustrates data points and trend lines to demonstrate matching to reduce model dependence.](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

• 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
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 ⇝ Model Dependence ⇝ Researcher discretion ⇝ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
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
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 $\leadsto$ 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[eff]

- conscientious effort doesn’t avoid biases (Banaji 2013)[acc]

- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)[exp]
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)
- 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 $\leadsto$ 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)

- 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:

\( \text{Imbalance} \leadsto \text{Model Dependence} \leadsto \text{Researcher discretion} \leadsto \text{Bias} \)
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)
- 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

- 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

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 $i$:

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

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

Quantities 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 $i$:

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

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

Quantities 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 $i$:
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 $i$:

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

= observed − unobserved
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]
- Estimate $Y_i(0)$ with $Y_j$ with 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 $Y_j$ with 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 \( Y_j \) with 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 $Y_j$ with a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[ SATT = \text{Mean } (TE_i) \]
        \[ i \in \{ T_i = 1 \} \]
  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 $Y_j$ with 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)
- **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 $i$:

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

$=$ observed $-$ unobserved

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

• Quantities 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
Types of Experiments

- Complete Randomization
- Fully Blocked
- Observed: On average
- Unobserved: On average

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

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

Types of Experiments

- Balance
  - Covariates:
    - Complete Randomization
    - Fully Blocked
  - Observed
  - 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

- Complete Randomization
- Fully Blocked
- Observational
- Unobserved

\[ \Rightarrow \text{Fully blocked dominates complete randomization for: imbalances, model dependence, power, efficiency, bias, research costs, robustness.} \]

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

Complete Randomization

⇝

Fully Blocked

On average

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

#### Types of Experiments

<table>
<thead>
<tr>
<th></th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>On average</td>
</tr>
<tr>
<td>Covariates</td>
<td>Complete</td>
<td>Fully Blocked</td>
</tr>
<tr>
<td>Balance</td>
<td>⇝</td>
<td></td>
</tr>
<tr>
<td>Goal of Each Matching Method (in Observational Data)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSM</td>
<td>Complete Randomization</td>
<td></td>
</tr>
<tr>
<td>Other methods</td>
<td>Fully Blocked</td>
<td></td>
</tr>
<tr>
<td>Other matching methods dominate PSM (wait, it gets worse)</td>
<td></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 Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given the structure of the table, it can be inferred that Fully Blocked randomization dominates Complete randomization for various criteria such as imbalance, model dependence, power, efficiency, bias, and research costs. For example, Imai, King, Nall (2009) report SEs that are 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></td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td></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></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

On average, fully blocked design 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></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>

Example: 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>

- 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>

$\Rightarrow$ *Fully blocked* dominates *complete randomization*
## Types of Experiments

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

$\Rightarrow$ **Fully blocked** dominates **complete randomization** for:
### 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>

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

\[\leadsto \text{Fully blocked dominates complete randomization for: imbalance, model dependence,}\]
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></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>

→ Fully blocked dominates complete randomization for:
imbalance, model dependence, power,
## 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>

According to Imai, King, Nall 2009, SEs are 600% smaller.

Goal of Each Matching Method (in Observational Data)

- **PSM**: complete randomization
- **Other methods**: fully blocked

Other matching methods dominate PSM.

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>

 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>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>

\[ \Rightarrow \text{Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs,} \]
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>

$\leadsto$ *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>[\text{On average}]</td>
<td>On average</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, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

| Balance | Complete |Fully |
| Covariates: | Randomization | Blocked |
| Observed | On average | Exact |
| Unobserved | On average | On average |

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

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></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>

\(\Rightarrow\) 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
### Types of Experiments

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

### 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>On average</td>
<td>On average</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!

#### 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>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** 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)
   - Distance($X_c, X_t$) = $\sqrt{\left(X_c - X_t\right)^\prime S^{-1} \left(X_c - X_t\right)}$
   - (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 $> \text{caliper}$
   - (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)
   - 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 $> \text{caliper}$

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

Educational Level (years)

12 14 16 18 20 22 24 26 28

20 30 40 50 60 70 80
Mahalanobis Distance Matching

Education (years)

Age

Education (years)

Age
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching

[Graph showing a scatter plot with Education (years) on the x-axis and Age on the y-axis. The data points are marked with 'C' and 'T' symbols.]
Method 2: Coarsened Exact Matching

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
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)
Method 2: Coarsened Exact Matching
(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
(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
(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
(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
(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
(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
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
(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

Age

Education

12 14 16 18 20 22 24 26 28

20
30
40
50
60
70
80

CCC
CC CC
C CC C
CCC CCCC CCCC
C CCC CC C

T T
T T
TT TT T TT
TTT TT
T T
T TT

11 / 28
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

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

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

Education

- HS
- BA
- MA
- PhD
- 2nd PhD

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

Drinking age

Education

HS BA MA PhD 2nd PhD

Senior Discounts

The Big 40

Don't trust anyone over 30

Drinking age
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching
Method 3: Propensity Score Matching

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
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)
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 $> \text{caliper}$

   (Many adjustments available to this basic method)

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

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
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

Education (years)

Age

12 16 20 24 28

20

30

40

50

60

70

80

C

C

CC

C

C

C

C

CC

C

CC

C

CC

C

C

C

C

CC

C

C

C

C

C

CC

CC

C

C

T TTT T

T

T

T

T

T

T

T

TT

T

T

T

T

T
Propensity Score Matching

Propensity Score

Education (years)

Age

12 16 20 24 28

20

30

40

50

60

70

80

C

C

CC

C

C

C

C

C

CC

C

CC

C

CC

C

C

C

C

CC

C

C

C

C

C

CC

CC

C

C

T TTT T

T

T

T

T

T

1

0

Propensity Score
Propensity Score Matching

Educational (years)
Age
12 16 20 24...
CC
C
C
C
C
C
C
C
CCCC
C
CCC
C
CC
CCC
C
C
C
C
C
C
T
TTT
T
TT
T
T
T
T
TT
T
T
T
T
T
T
T
Propensity Score Matching

Education (years)

Age

12 16 20 24 ...

Propensity Score

Education (years)

Propensity Score
Propensity Score Matching

Age

Education (years)

Propensity Score
Propensity Score Matching

![Graph showing the relationship between age and education years, with propensity scores indicated by 'C' and 'T'.]
Propensity Score Matching

Education (years)

Age

12 16 20 24 28
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

Age

Education (years)

Propensity Score
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching is Suboptimal
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 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)
       \[ \hat{\pi}_c \approx \pi_t \]
       \[ \Rightarrow \]
       Imbalance ➝ Inefficency ➝ Model dependence ➝ 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
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 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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \]

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)
     \[ \hat{\pi} \approx 0 \text{ (or constant within strata)} \]
     \[ \implies \text{pruning at random} \implies \text{Imbalance} \implies \text{Inefficency} \implies \text{Model dependence} \implies \text{Bias} \]
   - 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?
     \[ \text{Nope.} \text{ The PSM Paradox gets worse with more covariates} \]
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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \nleftrightarrow X_c = X_t \]
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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \]
     \[ \pi_c = \pi_t \implies X_c = X_t \]

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 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
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 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)
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 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)
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 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) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) 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 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) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) 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 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 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
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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \quad \not\implies \quad X_c = X_t \]

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
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 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) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) Imbalance \( \sim \) Inefficency \( \sim \) Model dependence \( \sim \) 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 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 \( \leadsto \) 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?
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 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 \( \leadsto \) Inefficiency \( \leadsto \) 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.
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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \neq \implies X_c = X_t \]

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 \( \rightsquigarrow \) Inefficiency \( \rightsquigarrow \) 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. 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

\[ Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i \]
\[ \epsilon_i \sim \mathcal{N}(0, 1) \]
The Propensity Score Paradox in Real Data

Similar pattern for >20 other real data sets we checked
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 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 ⇝ 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.
• 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 $\leadsto$ Imbalance-$n$ Trade Off
The Matching Frontier

- Frontier = matched dataset with lowest imbalance for each $n$
- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
- Simple to use
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
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
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
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 $\sim$ 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
- 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 $n = N, N-1, \ldots, 1$

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 $X_0$

\[
\binom{N}{n} \text{ evaluations for each sample size } n = N, N-1, \ldots, 1
\]

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 $X_0$

\[\binom{N}{n} \text{ evaluations for each sample size } n = N, N-1, \ldots, 1\]

- The combination is the (gargantuan) “power set”
- e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

\[\Rightarrow \text{It's hard to calculate!}\]

- 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)

\[\Rightarrow \text{It's easy to calculate!}\]
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
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 \( n = N, N-1, ..., 1 \)
  • 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 $n = N, N - 1, \ldots, 1$
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”

- 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)

- 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!
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
  - $\rightsquigarrow$ 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 $\binom{N}{n}$ 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
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
  - 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
Constructing the FSATT Mahalanobis Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2
-1.0 -0.5 0.0 0.5 1.0
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
Treated
Control
Next to remove

Frontier

Number of Observations Dropped

Covariate 1

Average Mahalanobis Discrepancy
0.0 0.1 0.2 0.3 0.4
●

0 5 10 15 20

24 / 28
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1
Covariate 2

Covariate 2
Covariate 1

Average Mahalanobis Discrepancy
Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1
−1.0 −0.5 0.0 0.5 1.0

Covariate 2
−1.0 −0.5 0.0 0.5 1.0

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Remaining Data

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

Treated
Control
Next to remove

Average Mahalanobis Discrepancy

0 0.1 0.2 0.3 0.4

Number of Observations Dropped

0 5 10 15 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2
−1.0 −0.5 0.0 0.5 1.0

●

●

Treated
Control
Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped

0.0
0.1
0.2
0.3
0.4

0 5 10 15 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

Graph showing the remaining data and the frontier with treated, control, and next to remove markers.
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1

Covariate 2

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0

-0.5

0.0

0.5

1.0

●

●

●

●...
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

Legend:
- Treated
- Control
- Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1
Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0
-0.5
0.0
0.5
1.0
●

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20
0.0
0.1
0.2
0.3
0.4 ●
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the FSATT Mahalanobis Frontier

**Remaining Data**

- Covariates 1 and 2
- Treated
- Control
- Next to remove

**Frontier**

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

- Remaining data points are plotted with different symbols representing the groups and the next to remove observations.
Constructing the FSATT Mahalanobis Frontier

**Remaining Data**
- Covariate 1
- Covariate 2

**Frontier**
- Treated
- Control
- Next to remove

**Average Mahalanobis Discrepancy**
- Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

<table>
<thead>
<tr>
<th>Number of Observations Dropped</th>
<th>Average Mahalanobis Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>0.2</td>
</tr>
<tr>
<td>15</td>
<td>0.1</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data
- Covariate 1
- Covariate 2

Frontier
- Number of Observations Dropped
- Average Mahalanobis Discrepancy

● Treated
○ Control
● Next to remove

• Warning: figure omits details and the proof!
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Warning: figure omits details and the proof!
- Very fast; works with any continuous imbalance metric
Constructing the L1/L2 SATT Frontier

![Graph showing frequency and number of observations dropped across different bins.](image-url)
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distribution in different bins for Treatment and Control groups.]

- Frequency
- Treatment
- Control

- Number of Observations Dropped
- L2

- Graph showing the number of observations dropped vs. L2 value.
Constructing the L1/L2 SATT Frontier

![Graph showing frequency distribution and L2 values across different bins.]

- Bins 1 to 6 with frequency counts.
- L2 values for Observation Dropped.

Number of Observations Dropped

- 25/28
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency of Treatment and Control in different bins.](image)

![Graph showing the number of observations dropped vs. L2 values.](image)
Constructing the L1/L2 SATT Frontier

![Bar Chart]

- **Bin 1**: Treatment 4, Control 4
- **Bin 2**: Treatment 6, Control 7
- **Bin 3**: Treatment 2, Control 3
- **Bin 4**: Treatment 3, Control 4
- **Bin 5**: Treatment 3, Control 2
- **Bin 6**: Treatment 1, Control 2

![Line Graph]

- **L2** vs. **Number of Observations Dropped**
- **L2** values: 0.12, 0.10, 0.08, 0.06, 0.04, 0.02
- **Number of Observations Dropped**: 0, 2, 4, 6, 8, 10

Legend:
- □ Treatment
- ▢ Control
Constructing the L1/L2 SATT Frontier

![Graph showing the distribution of observations in different bins.](image)

- **Bin 1:** Treatment: 4, Control: 6
- **Bin 2:** Treatment: 6, Control: 6
- **Bin 3:** Treatment: 2, Control: 3
- **Bin 4:** Treatment: 3, Control: 4
- **Bin 5:** Treatment: 3, Control: 2
- **Bin 6:** Treatment: 1, Control: 2

![Graph showing the relationship between the number of observations dropped and L2 value.](image)

- **Number of Observations Dropped:** 0, 2, 4, 6, 8, 10
- **L2 Values:** 0.12, 0.10, 0.08, 0.06, 0.04, 0.02, 0.00
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distributions for different bins and a line graph showing the number of observations dropped.](chart.png)
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequencies in bins 1 to 6 for Treatment and Control groups.](chart)

![Graph showing the L2 value decreasing with the number of observations dropped.](graph)

- **Frequency Distribution**: Bar chart with bins labeled Bin1 to Bin6, showing the number of observations in each bin for Treatment and Control groups.
- **Graph**: Graph with the x-axis labeled "Number of Observations Dropped" and the y-axis labeled "L2". The graph shows a downward trend as the number of observations dropped increases.
Constructing the L1/L2 SATT Frontier

Frequency

<table>
<thead>
<tr>
<th>Bin</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Bin2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Bin3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bin4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bin5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bin6</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Number of Observations Dropped

L2

0.00  0.02  0.04  0.06  0.08  0.10  0.12

0  2  4  6  8  10

Number of Observations Dropped
Constructing the L1/L2 SATT Frontier

Frequency

<table>
<thead>
<tr>
<th>Bin</th>
<th>Frequency</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Bin2</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Bin3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bin4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Bin5</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bin6</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Number of Observations Dropped

- L2 graph
  - L2 values: 0.12, 0.10, 0.08, 0.06, 0.04, 0.02, 0.00
  - X-axis: Number of Observations Dropped from 0 to 10
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
- Works very fast, even with very large data sets
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;
  - 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;
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
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