Detecting and Reducing Model Dependence in Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

Peking University, 1/26/2017

\(^1\)GaryKing.org
Readings, Part 1: Model Dependence


Related Software: WhatIf, MatchIt, Zelig, CEM

http://j.mp/causalinference
Readings, Part 1: Model Dependence

Readings, Part 1: Model Dependence

Readings, Part 1: Model Dependence

- Related Software: WhatIf, MatchIt, Zelig, CEM
Readings, Part 1: Model Dependence


• Related Software: WhatIf, MatchIt, Zelig, CEM

http://j.mp/causalinferece
Counterfactuals

Three types:
1. Forecasts: What will the mortality rate be in 2025?
2. What-if Questions: What would have happened if the U.S. had not invaded Iraq?
3. Causal Effects: What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)
Counterfactuals

• Three types:

1. Forecasts: What will the mortality rate be in 2025?
2. What-if Questions: What would have happened if the U.S. had not invaded Iraq?
3. Causal Effects: What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)
Counterfactuals

- Three types:
  1. **Forecasts** What will the mortality rate be in 2025?
  2. **What if** questions What would have happened if the U.S. had not invaded Iraq?
  3. Causal effects What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)
Counterfactuals

- Three types:
  1. Forecasts What will the mortality rate be in 2025?
  2. Whatif Questions What would have happened if the U.S. had not invaded Iraq?
Counterfactuals

• Three types:
  1. **Forecasts** What will the mortality rate be in 2025?
  2. **Whatif Questions** What would have happened if the U.S. had not invaded Iraq?
  3. **Causal Effects** What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)
Counterfactuals

• Three types:
  1. Forecasts What will the mortality rate be in 2025?
  2. What if Questions What would have happened if the U.S. had not invaded Iraq?
  3. Causal Effects What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)

• Counterfactuals are part of most social science research
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$.
- How do you choose a model? $R^2$, some "test", "theory"?

The bottom line: answers to some questions don't exist in the data. We show how to determine which ones.

Same for what if questions, predictions, and causal inferences.
• Compare prediction at $x = 1.5$ to prediction at $x = 5$
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model?
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? $R^2$?
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? $R^2$? Some “test”? 

![Graph showing comparison of linear and quadratic models]
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$
Which model would you choose? (Both fit the data well.)

- Compare prediction at \( x = 1.5 \) to prediction at \( x = 5 \)
- How do you choose a model? \( R^2 \)? Some “test”? “Theory”?
- The bottom line: answers to some questions don’t exist in the data. We show how to determine which ones.
Which model would you choose? (Both fit the data well.)

- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- The bottom line: answers to some questions don’t exist in the data. We show how to determine which ones.
- Same for what if questions, predictions, and causal inferences
Model Dependence Proof

To estimate $E(Y|X=x)$ at $x$, average many observed $Y$ with value $x$.

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive).

Result

The maximum degree of model dependence: a function of the distance from the counterfactual to the data.
Model Dependence Proof

Model Free Inference
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X = x)$ at $x$, average many observed $Y$ with value $x$. 

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive).

Result
The maximum degree of model dependence: a function of the distance from the counterfactual to the data.
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X = x)$ at $x$, average many observed $Y$ with value $x$

Assumptions (Model-Based Inference)
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X = x)$ at $x$, average many observed $Y$ with value $x$.

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X=x)$ at $x$, average many observed $Y$ with value $x$

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.

2. The functional form follows strong continuity (think smoothness, although it is less restrictive)
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X = x)$ at $x$, average many observed $Y$ with value $x$

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.

2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Result
Model Dependence Proof

Model Free Inference
To estimate $E(Y|X = x)$ at $x$, average many observed $Y$ with value $x$

Assumptions (Model-Based Inference)

1. Definition: model dependence at $x$ is the difference between predicted outcomes for any two models that fit about equally well.

2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Result
The maximum degree of model dependence: a function of the distance from the counterfactual to the data
A Simple Measure of Distance from The Data

Figure: The Convex Hull
A Simple Measure of Distance from The Data

Figure: The Convex Hull

- **Interpolation**: Inside the convex hull
A Simple Measure of Distance from The Data

Figure: The Convex Hull

- **Interpolation**: Inside the convex hull
- **Extrapolation**: Outside the convex hull
A Simple Measure of Distance from The Data

- **Interpolation**: Inside the convex hull
- **Extrapolation**: Outside the convex hull
- Works mathematically for any number of $X$ variables

Figure: The Convex Hull
Interpolation: Inside the convex hull
Extrapolation: Outside the convex hull
Works mathematically for any number of \( X \) variables
Software to determine whether a point is in the hull (which is all we need) without calculating the hull (which would take forever), so its fast; see GaryKing.org/whatif
Model Dependence Example

• Data: 124 Post-World War II civil wars
• Dependent var: peacebuilding success
• Treatment: multilateral UN peacekeeping intervention (0/1)
• Control vars: war type, severity, duration; development status, ...
• Counterfactual question: Switch UN intervention for each war
• Data analysis: Logit model
• The question: How model dependent are the results?
• Percent of counterfactuals in the convex hull: 0%
• ⇝ without estimating any models, we know: inferences will be model dependent
• For illustration: let's find an example...
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

• Data: 124 Post-World War II civil wars
• Dependent var: peacebuilding success
• Treatment: multilateral UN peacekeeping intervention (0/1)
• Control vars: war type, severity, duration; development
• Counterfactual question: Switch UN intervention for each war
• Data analysis: Logit model
• The question: How model dependent are the results?

\[
\text{Percent of counterfactuals in the convex hull: } 0\%
\]

\[
\Rightarrow \text{without estimating any models, we know: inferences will be model dependent}
\]

For illustration: let's find an example...
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data**: 124 Post-World War II civil wars
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data:** 124 Post-World War II civil wars
- **Dependent var:** peacebuilding success
Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data:** 124 Post-World War II civil wars
- **Dependent var:** peacebuilding success
- **Treatment:** multilateral UN peacekeeping intervention (0/1)
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data**: 124 Post-World War II civil wars
- **Dependent var**: peacebuilding success
- **Treatment**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status, ...
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data:** 124 Post-World War II civil wars
- **Dependent var:** peacebuilding success
- **Treatment:** multilateral UN peacekeeping intervention (0/1)
- **Control vars:** war type, severity, duration; development status, ...
- **Counterfactual question:** Switch UN intervention for each war
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

• Data: 124 Post-World War II civil wars
• Dependent var: peacebuilding success
• Treatment: multilateral UN peacekeeping intervention (0/1)
• Control vars: war type, severity, duration; development status, ...
• Counterfactual question: Switch UN intervention for each war
• Data analysis: Logit model
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data**: 124 Post-World War II civil wars
- **Dependent var**: peacebuilding success
- **Treatment**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status,…
- **Counterfactual question**: Switch UN intervention for each war
- **Data analysis**: Logit model
- **The question**: How *model dependent* are the results?
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data**: 124 Post-World War II civil wars
- **Dependent var**: peacebuilding success
- **Treatment**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status, ...
- **Counterfactual question**: Switch UN intervention for each war
- **Data analysis**: Logit model
- **The question**: How *model dependent* are the results?
- **Percent of counterfactuals in the convex hull**: 0%
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data**: 124 Post-World War II civil wars
- **Dependent var**: peacebuilding success
- **Treatment**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status, ...
- **Counterfactual question**: Switch UN intervention for each war
- **Data analysis**: Logit model
- **The question**: How *model dependent* are the results?
- **Percent of counterfactuals in the convex hull**: 0%
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data:** 124 Post-World War II civil wars
- **Dependent var:** peacebuilding success
- **Treatment:** multilateral UN peacekeeping intervention (0/1)
- **Control vars:** war type, severity, duration; development status, ...
- **Counterfactual question:** Switch UN intervention for each war
- **Data analysis:** Logit model
- **The question:** How *model dependent* are the results?
- **Percent of counterfactuals in the convex hull:** 0%
- **∼ without estimating any models, we know:** inferences will be model dependent
Model Dependence Example
Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- **Data:** 124 Post-World War II civil wars
- **Dependent var:** peacebuilding success
- **Treatment:** multilateral UN peacekeeping intervention (0/1)
- **Control vars:** war type, severity, duration; development status,
- **Counterfactual question:** Switch UN intervention for each war
- **Data analysis:** Logit model
- **The question:** How *model dependent* are the results?
- **Percent of counterfactuals in the convex hull:** 0%
- **⇝** without estimating any models, we know: inferences will be model dependent
- **For illustration:** let’s find an example...
Two Logit Models, Apparently Similar Results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff</th>
<th>SE</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wartype</td>
<td>-1</td>
<td>0.742</td>
<td>0.004</td>
</tr>
<tr>
<td>Logdead</td>
<td>-0.445</td>
<td>0.126</td>
<td>0.000</td>
</tr>
<tr>
<td>Wardur</td>
<td>0.006</td>
<td>0.006</td>
<td>0.258</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1</td>
<td>0.259</td>
<td>0.073</td>
</tr>
<tr>
<td>Factnum2</td>
<td>0.062</td>
<td>0.065</td>
<td>0.346</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>0.004</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>Develop</td>
<td>0.001</td>
<td>0.000</td>
<td>0.065</td>
</tr>
<tr>
<td>Exp</td>
<td>-6</td>
<td>0.016</td>
<td>0.050</td>
</tr>
<tr>
<td>Decade</td>
<td>-0.299</td>
<td>0.169</td>
<td>0.077</td>
</tr>
<tr>
<td>Treaty</td>
<td>2</td>
<td>1.124</td>
<td>0.010</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3</td>
<td>0.135</td>
<td>0.004</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
<td>0.001</td>
</tr>
<tr>
<td>Constant</td>
<td>8</td>
<td>2.609</td>
<td>0.000</td>
</tr>
</tbody>
</table>

N: 122
Log-likelihood: -45.649, -44.902
Pseudo R^2: 0.423, 0.433
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff</th>
<th>SE</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wartype</td>
<td>−1.742</td>
<td>0.609</td>
<td>0.004</td>
</tr>
<tr>
<td>Logdead</td>
<td>−0.445</td>
<td>0.126</td>
<td>0.000</td>
</tr>
<tr>
<td>Wardur</td>
<td>0.006</td>
<td>0.006</td>
<td>0.258</td>
</tr>
<tr>
<td>Factnum</td>
<td>−1.259</td>
<td>0.703</td>
<td>0.073</td>
</tr>
<tr>
<td>Factnum2</td>
<td>0.062</td>
<td>0.065</td>
<td>0.346</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>0.004</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>Develop</td>
<td>0.001</td>
<td>0.000</td>
<td>0.065</td>
</tr>
<tr>
<td>Exp</td>
<td>−6.016</td>
<td>3.071</td>
<td>0.050</td>
</tr>
<tr>
<td>Decade</td>
<td>−2.99</td>
<td>1.169</td>
<td>0.077</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>0.821</td>
<td>0.010</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
<td>0.004</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—as—</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609</td>
<td>2.157</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Log-likelihood: -45.649, -44.902
Pseudo R²: .423, .433
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

<table>
<thead>
<tr>
<th>Variables</th>
<th>Original “Interactive” Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>SE</td>
</tr>
<tr>
<td>Wartype</td>
<td>-1.742</td>
<td>.609</td>
</tr>
<tr>
<td>Logdead</td>
<td>- .445</td>
<td>.126</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.259</td>
<td>.703</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.062</td>
<td>.065</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.016</td>
<td>3.071</td>
</tr>
<tr>
<td>Decade</td>
<td>- .299</td>
<td>.169</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609</td>
<td>2.157</td>
</tr>
</tbody>
</table>

N              | 122    | 122    |
Log-likelihood  | -45.649 | -44.902 |
Pseudo R²       | .423   | .433   |
Model Dependence: Same Fit, Different Predictions

In Sample Fit

Counterfactual Prediction
Readings, Part 2: Matching

1. Do powerful methods have to be complicated?

   "Causal Inference Without Balance Checking: Coarsened Exact Matching" (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

2. The most popular method (propensity score matching, used in 76,900 articles!) sounds magical:

   "Why Propensity Scores Should Not Be Used for Matching" (Gary King, Richard Nielsen)

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)
Readings, Part 2: Matching

1. Do powerful methods have to be complicated?

[Citation: Coarsened Exact Matching]

[Citation: Why Propensity Scores Should Not Be Used for Matching]

[Citation: The Balance-Sample Size Frontier in Matching Methods for Causal Inference]
1. Do powerful methods have to be complicated?
   ⇒ “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)
Readings, Part 2: Matching

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

2. The most popular method (propensity score matching, used in 76,900 articles!) sounds magical:
Readings, Part 2: Matching

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

2. The most popular method (propensity score matching, used in 76,900 articles!) sounds magical:
   ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)
Readings, Part 2: Matching

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

2. The most popular method (propensity score matching, used in 76,900 articles!) sounds magical:
   ~ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

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

Readings, Part 2: Matching

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

2. The most popular method (propensity score matching, used in 76,900 articles!) sounds magical:
   ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

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*)

Education (years)

<table>
<thead>
<tr>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

Education (years)
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)
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 $\sim$ Model Dependence
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ 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
conscienous 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 $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ 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
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 $\sim$ 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)
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 $\Rightarrow$ 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)
- “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 $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance $\sim$ Model Dependence $\sim$ Researcher discretion $\sim$ 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?

- $Y_i$ (dependent variable), $T_i$ (treatment indicator, $1=$ treated, $0=$ control), and $X_i$ (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
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 $i$:

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

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

= 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:

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

  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:

\[ \text{SATT} = \text{Mean}_i \{ TE_i \}_{i \in \{ T_i = 1 \}} \]

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

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

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

  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
Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- \( n = 408 \) new drugs (262 approved, 146 pending)
- Lognormal survival model
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)
Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
Empirical Illustration: Carpenter, AJPS, 2002

• Hypothesis: Democratic senate majorities slow FDA drug approval time

• $n = 408$ new drugs (262 approved, 146 pending).
Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- \( n = 408 \) new drugs (262 approved, 146 pending).
- lognormal survival model.
Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- \( n = 408 \) new drugs (262 approved, 146 pending).
- lognormal survival model.
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)
Evaluating Reduction in Model Dependence

• Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
• Match: prune 49 units (2 treated, 17 control units).
• run 262,143 possible specifications and calculates SATT for each.
• Look at variability in SATT estimate across specifications.
• (Normal applications would only use one or a few specifications.)
Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
- run 262,143 possible specifications and calculates SATT for each.
• Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
• Match: prune 49 units (2 treated, 17 control units).
• run 262,143 possible specifications and calculates SATT for each.
• Look at variability in SATT estimate across specifications.
Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
- run 262,143 possible specifications and calculates SATT for each.
- Look at \textit{variability} in SATT estimate across specifications.
- (Normal applications would only use one or a few specifications.)
Reducing Model Dependence

Figure: SATT Histogram: Effect of Democratic Senate majority on FDA drug approval time, across 262,143 specifications.
Another Example: Jeffrey Koch, AJPS, 2002

Figure: SATT Histogram: Effect of being a highly visible female Republican candidate across 63 possible specifications with the Koch data.
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

- Balance
  - Covariates:
    - 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!
### Types of Experiments

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

- Observed: On average
- Unobserved: On average

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

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 (wait, it gets worse)
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

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

- **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

<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>

On average, fully blocked randomization 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>On average</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</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</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 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></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 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>

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

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

- 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

<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,
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>

\[ \therefore \text{Fully blocked dominates complete randomization for:} \]
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></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></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\rightsquigarrow \text{Fully blocked dominates complete randomization for:}
\text{imbalance, model dependence, power, efficiency,}\]
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>Unobserved</td>
<td></td>
</tr>
</tbody>
</table>

≈ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias,
### 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>

~> **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></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: imbalance, model dependence, power, efficiency, bias, research costs, robustness.
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>

⇝ 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></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></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
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
- Other methods: fully blocked
Matching: Finding Hidden Randomized Experiments

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>

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

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></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

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

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)
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**

2. **Estimation** Difference in means or a model

\[ \text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)} \]

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

Education (years)

Age
12 14 16 18 20 22 24 26 28

20
30
40
50
60
70
80

T TTT T
T
T
T
T
T
T
T
TT
T
T
T
T
T
Mahalanobis Distance Matching

Age

12 14 16 18 20 22 24 26 28

Education (years)
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching

![Graph showing data points related to age and education (years).]
Best Case: Mahalanobis Distance Matching
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)

1. Preprocess (Matching)
   • Temporarily coarsen \( \mathbf{X} \) as much as you're willing
   • e.g., Education (grade school, high school, college, graduate)
   • Apply exact matching to the coarsened \( \mathbf{X} \), \( \mathbf{C}(\mathbf{X}) \)
   • Sort observations into strata, each with unique values of \( \mathbf{C}(\mathbf{X}) \)
   • 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)

1. **Preprocess** (Matching)

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
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
     - 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 vs. Age

Age

Education
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

Education

HS | BA | MA | PhD | 2nd PhD
Coarsened Exact Matching

Education

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

Education

HS BA MA PhD 2nd PhD
Drinking age

C C C
CC C CC
C C
C
TT T T TT
TTT TT
T T
Coarsened Exact Matching

![Graph showing a scatter plot with education on the y-axis and age on the x-axis. The points are marked 'C' or 'T' and are distributed across the range of values.](image-url)
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 $> \text{caliper}$
   - (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 $\pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}$
   - 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 $> \text{caliper}$

2. Estimation
   - Difference in means or a model
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
     \[ \pi_i \equiv \Pr(T_i = 1 \mid X) = \frac{1}{1 + e^{-x_i \beta}} \]

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

Education (years)
Propensity Score Matching

Education (years) vs. Age

Propensity Score
Propensity Score Matching

Age
12 16 20 24 28

Education (years)

Propensity Score

20
30
40
50
60
70
80

0
1

0
1

26 / 41
Propensity Score Matching

Education (years) vs. Age

- Age: 12, 16, 20, 24, 28
- Education (years): 20, 30, 40, 50, 60, 70, 80

C and T symbols indicate different groups or conditions.
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

![Graph showing the relationship between age and education years, with a focus on propensity score matching.](image)
Best Case: Propensity Score Matching

![Graph showing the relationship between age and education](image)
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: $\{M_t, F_c\}$, $\{F_t, M_c\}$.

- \[\Rightarrow\text{random pruning increases imbalance}\]

- Continuous example:
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units.
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$.
  - $E(d^2) = \frac{\sigma^2}{n}$ (note: $E(d) = 0$).

- \[\Rightarrow\text{random pruning increases imbalance}\]

- 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

• Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$

• $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)

• 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 $X$

$\Rightarrow$ 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
- Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
- $E(d^2) = V(d) \propto \frac{1}{n}$ (note: $E(d) = 0$)

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
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$

  $\mathbb{E}(d^2) = \mathbb{V}(d) \propto \frac{1}{n}$ (note: $\mathbb{E}(d) = 0$)

• Random pruning $\Rightarrow n$ declines $\Rightarrow \mathbb{E}(d^2)$ 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 $\leftrightarrow$ 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 $\rightsquigarrow$ 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
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
  - $\mathbb{E}(d^2) = \mathbb{V}(d) \propto 1/n$ (note: $\mathbb{E}(d) = 0$)
  - Random pruning $\leadsto n$ declines $\leadsto \mathbb{E}(d^2)$ 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: treated $M_t$, $F_t$, 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
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
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 $\leadsto$ 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
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
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 $\rightsquigarrow$ 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
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow n$ declines $\rightsquigarrow E(d^2)$ 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
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow n$ declines $\rightsquigarrow E(d^2)$ increases
  - $\implies$ 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 \( \rightsquigarrow \) 4 possible datasets:
    2 balanced \( \{ M_t, M_c \}, \{ F_t, F_c \} \)
    2 imbalanced \( \{ M_t, F_c \}, \{ F_t, M_c \} \)
  - \( \Longrightarrow \) random pruning increases imbalance

• **Continuous example**
  - Dataset: \( T \in \{0, 1\} \) randomly assigned; \( X \) any fixed variable; with \( n \) units
  - Measure of imbalance: squared difference in means \( d^2 \), where \( d = \bar{X}_t - \bar{X}_c \)
  - \( E(d^2) = V(d) \propto 1/n \) (note: \( E(d) = 0 \))
  - Random pruning \( \rightsquigarrow n \) declines \( \rightsquigarrow E(d^2) \) increases
  - \( \Longrightarrow \) 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

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) ⇝ \( \hat{\pi}_c \approx \pi_c \) within strata ⇝ pruning at random ⇝ 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

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)
   - \( \hat{\pi} \approx 0 \) (or constant within strata)
   - \( \Rightarrow \) pruning at random
   - \( \Rightarrow \) Imbalance
   - \( \Rightarrow \) Inefficency
   - \( \Rightarrow \) Model dependence
   - \( \Rightarrow \) 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?
     - No.
     - 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
1. **Low Standards**: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \]
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:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]
1. **Low Standards**: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \] but
     \[ \pi_c = \pi_t \implies X_c = X_t \]

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:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox: When you do “better,” you do worse**
   - Background: Random matching increases imbalance
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)
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)
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) \( \implies \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \implies \) 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:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

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

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
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) \( \implies \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \implies \) pruning at random \( \implies \) Imbalance \( \implies \) Inefficency \( \implies \) 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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \implies X_c = X_t \]

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.
   - 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:
     \[ X_c = X_t \implies \pi_c = \pi_t \quad \text{but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

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) \( \rightsquigarrow \) 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 dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \] 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) \( \forall \hat{\pi} \approx 0.5 \) (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.
   - 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:  \(X_1, X_2 \sim \text{Uniform}(0,5)\)
Treateds:  \(X_1, X_2 \sim \text{Uniform}(1,6)\)
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

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) and Nielsen et al. (AJPS, 2011) show similar patterns for > 20 other real data sets we checked.

- For Finkel et al.: Imbalance decreases as the number of units pruned increases. The plots show that CEM, MDM, and PSM have lower imbalance compared to random assignment.
- For Nielsen et al.: Similar trends are observed, with CEM and MDM performing better than random assignment.

These findings highlight the importance of careful intervention selection and the potential paradoxes that can arise in real-world data.
The Matching Frontier

• Frontier = matched dataset with lowest imbalance for each $n$

• 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 $\leadsto$ 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 $\rightsquigarrow$ 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 $\leadsto$ 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 $\rightsquigarrow$ 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 $n = N, N-1, \ldots, 1$
- 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:
How hard is the frontier to calculate?

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

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

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

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
- 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 \( n = N, N - 1, \ldots, 1 \)

\( \Rightarrow \) 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

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

- 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)

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

- 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
  - $\Rightarrow$ 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 $\binom{N}{n}$ 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 \( \binom{N}{n} \) 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)
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
  - $\leadsto$ 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)
  - $\leadsto$ It’s easy to calculate!
• 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

- Treated
- Control
- Next to remove

**Frontier**

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

Remaining Data

- Covariate 1
- Covariate 2

Covariate 1

-1.0 −0.5 0.0 0.5 1.0

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

-1.0 -0.5 0.0 0.5 1.0

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

- Covariate 1
- Covariate 2

Frontier

- 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
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy

Covariate 1

Covariate 2

Treated
Control
Next to remove
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

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped
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

● Treated
● Control
● Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

0
0.1
0.2
0.3
0.4
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
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

Number of Observations Dropped

0 5 10 15 20

Average Mahalanobis Discrepancy

0.0 0.1 0.2 0.3 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

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

Frontier

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
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

●

●

●

●

●

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

0.0

0.1

0.2

0.3

0.4

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
−1.0
−0.5
0.0
0.5
1.0
●
●
●
●
●
Treated
Control
Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

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

Remaining Data

- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

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

![Bar chart and distribution graph showing frequency and L2 values for different bins.](image)
Constructing the L1/L2 SATT Frontier

![Frequency vs. Number of Observations Dropped](image)

- **Frequency:**
  - Bin1: Control (4), Treatment (5)
  - Bin2: Control (6), Treatment (7)
  - Bin3: Control (2), Treatment (3)
  - Bin4: Control (3), Treatment (6)
  - Bin5: Control (3), Treatment (2)
  - Bin6: Control (2), Treatment (1)

- **Number of Observations Dropped:**
  - L2: 38/41
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

![Bar chart and line graph showing frequency and L2 values across bins with treatment and control groups.]

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

- Line graph showing L2 values decreasing with increasing number of observations dropped.

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]

- **Frequency**
- **Bins:** Bin1, Bin2, Bin3, Bin4, Bin5, Bin6
- **Comparison:** Treatment vs. Control

![Graph]

- **Y-Axis:** L2
- **X-Axis:** Number of Observations Dropped
- **Data Points:**
  - L2 value decreases as the number of observations dropped increases.
Constructing the L1/L2 SATT Frontier

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

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

- **L2 Values**
  - Number of observations dropped vs. L2 values
  - L2 values decrease as the number of observations dropped increases.
Constructing the L1/L2 SATT Frontier

Number of Observations Dropped

Treatment vs. Control

Frequency
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distribution across bins for Treatment and Control groups.](chart1.png)

![Graph showing L2 values against the number of observations dropped.](chart2.png)
Constructing the L1/L2 SATT Frontier

![Graph showing frequency distribution across different bins and a plot of L2 values against number of observations dropped.]

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

- **L2 Values**:
  - Number of Observations Dropped vs. L2 Value
  - L2 values range from 0.12 to 0.00
Constructing the L1/L2 SATT Frontier

- **Frequency Distribution**
  - Bin 1: 4 (Treatment), 4 (Control)
  - Bin 2: 6 (Treatment), 6 (Control)
  - Bin 3: 2 (Treatment), 2 (Control)
  - Bin 4: 3 (Treatment), 3 (Control)
  - Bin 5: 2 (Treatment), 2 (Control)
  - Bin 6: 1 (Treatment), 1 (Control)

- **Number of Observations Dropped**
  - L2: 38 / 41

- **Graph**
  - X-axis: Number of Observations Dropped
  - Y-axis: L2
  - Data points decrease as the number of observations dropped increases.
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

- 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

- 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
  - 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;

- 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;
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