Advanced Quantitative Research Methodology, Lecture Notes: Detecting and Reducing Model Dependence in Causal Inference

Gary King
Institute for Quantitative Social Science
Harvard University

April 14, 2019

¹GaryKing.org
Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier
Readings in Model Dependence


• Related Software: WhatIf, MatchIt, Zelig, CEM

http://j.mp/causalinference
Readings in Model Dependence

Readings in Model Dependence

Readings in Model Dependence


• Related Software: WhatIf, MatchIt, Zelig, CEM
Readings in Model Dependence


• Related Software: WhatIf, MatchIt, Zelig, CEM

http://j.mp/causalinference
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.
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.
Counterfactuals

• Three types:
  1. **Forecasts**  What will the mortality rate be in 2025?

• Counterfactuals are part of most social science research.
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?

[9x251]Counterfactuals

Detecting Model Dependence 4 / 44
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?
  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?
- $R^2$?
- "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.
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?
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”?
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”?
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.
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
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)
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 $x$
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

- 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 it's fast; see GaryKing.org/whatif
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

Figure: The Convex Hull

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

Detecting Model Dependence
A Simple Measure of Distance from The Data

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](http://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 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
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:
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%
- \( \Rightarrow \) *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>Variable</th>
<th>Original &quot;Interactive&quot; Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wartype</td>
<td>−1.742 0.609 0.004</td>
<td>−1.666 0.606 0.006</td>
</tr>
<tr>
<td>Logdead</td>
<td>−0.445 0.126 0.000</td>
<td>−0.437 0.125 0.000</td>
</tr>
<tr>
<td>Wardur</td>
<td>0.006 0.006 0.258</td>
<td>0.006 0.006 0.342</td>
</tr>
<tr>
<td>Factnum</td>
<td>−1.259 0.703 0.073</td>
<td>−1.045 0.899 0.245</td>
</tr>
<tr>
<td>Factnum2</td>
<td>0.062 0.065 0.346</td>
<td>0.032 0.104 0.756</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>0.004 0.002 0.010</td>
<td>0.004 0.002 0.017</td>
</tr>
<tr>
<td>Develop</td>
<td>0.001 0.000 0.065</td>
<td>0.001 0.000 0.068</td>
</tr>
<tr>
<td>Exp</td>
<td>−6.016 3.071 0.050</td>
<td>−6.215 3.065 0.043</td>
</tr>
<tr>
<td>Decade</td>
<td>−0.299 0.169 0.077</td>
<td>−0.284 0.169 0.093</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124 0.821 0.010</td>
<td>2.126 0.802 0.008</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135 1.091 0.004</td>
<td>0.262 1.392 0.851</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>0.037 0.011 0.001</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609 2.157 0.000</td>
<td>7.978 2.350 0.000</td>
</tr>
<tr>
<td>N</td>
<td>122</td>
<td>122</td>
</tr>
</tbody>
</table>

Log-likelihood: -45.649 vs. -44.902

Pseudo $R^2$: 0.423 vs. 0.433

Detecting Model Dependence
## Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

<table>
<thead>
<tr>
<th>Variable</th>
<th>Modified Model Coeff</th>
<th>SE</th>
<th>P-val</th>
<th>Original “Interactive” Model Coeff</th>
<th>SE</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wartype</td>
<td>-1.666</td>
<td>.606</td>
<td>.006</td>
<td>-1.742</td>
<td>.609</td>
<td>.004</td>
</tr>
<tr>
<td>Logdead</td>
<td>-0.437</td>
<td>.125</td>
<td>.000</td>
<td>-0.445</td>
<td>.126</td>
<td>.000</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
<td>.342</td>
<td>.006</td>
<td>.006</td>
<td>.258</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.045</td>
<td>.899</td>
<td>.245</td>
<td>-1.259</td>
<td>.703</td>
<td>.073</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.032</td>
<td>.104</td>
<td>.756</td>
<td>.062</td>
<td>.065</td>
<td>.346</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
<td>.017</td>
<td>.004</td>
<td>.002</td>
<td>.010</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
<td>.068</td>
<td>.001</td>
<td>.000</td>
<td>.065</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.215</td>
<td>3.065</td>
<td>.043</td>
<td>-6.016</td>
<td>3.071</td>
<td>.050</td>
</tr>
<tr>
<td>Decade</td>
<td>-0.284</td>
<td>.169</td>
<td>.093</td>
<td>-0.299</td>
<td>.169</td>
<td>.077</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.126</td>
<td>.802</td>
<td>.008</td>
<td>2.124</td>
<td>.821</td>
<td>.010</td>
</tr>
<tr>
<td>UNOP4</td>
<td>0.262</td>
<td>1.392</td>
<td>.851</td>
<td>3.135</td>
<td>1.091</td>
<td>.004</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td></td>
<td></td>
<td>.037</td>
<td>.011</td>
<td>.001</td>
</tr>
<tr>
<td>Constant</td>
<td>7.978</td>
<td>2.350</td>
<td>.000</td>
<td>8.609</td>
<td>2.157</td>
<td>.000</td>
</tr>
</tbody>
</table>

N = 122

Log-likelihood = -44.902

Pseudo \(R^2\) = .433

Detecting Model Dependence
## 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>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>
<tr>
<td>N</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−45.649</td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.423</td>
<td></td>
</tr>
</tbody>
</table>

Detecting Model Dependence
Model Dependence: Same Fit, Different Predictions

In Sample Fit

Counterfactual Prediction

Detecting Model Dependence
Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier
Readings, Matching

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

• The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ↝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• 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, 2017; Gary King, Christopher Lucas and Richard Nielsen)

• Current practice, matching as preprocessing: violates current statistical theory.
  So let’s change the theory:
  ↝ “A Theory of Statistical Inference for Matching Methods in Causal Research” (Stefano Iacus, Gary King, Giuseppe Porro)
Readings, Matching

- Do powerful methods have to be complicated?
Readings, Matching

- 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, Matching

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

• The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
Readings, Matching

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

- The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  - “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)
Readings, Matching

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

• The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ↝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously’:
Readings, Matching

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

- The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ↗  “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

- 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, 2017; Gary King, Christopher Lucas and Richard Nielsen)
Readings, Matching

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

• The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ↗️ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• 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, 2017; Gary King, Christopher Lucas and Richard Nielsen)

• Current practice, matching as preprocessing:
Readings, Matching

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

- The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ~ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

- 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, 2017; Gary King, Christopher Lucas and Richard Nielsen)

- Current practice, matching as preprocessing: violates current statistical theory.
Readings, Matching

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

• The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  ↝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• 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, 2017; Gary King, Christopher Lucas and Richard Nielsen)

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
Readings, Matching

- Do powerful methods have to be complicated?

- The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  - “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

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

- Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  - “A Theory of Statistical Inference for Matching Methods in Causal Research” (Stefano Iacus, Gary King, Giuseppe Porro)
Matching to Reduce Model Dependence
Matching to Reduce Model Dependence

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

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

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

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

Education (years)

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

C
T

12
10
8
6
4
2
0

Education (years)

12 14 16 18 20 22 24 26 28
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)
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

Without matching, qualitative choice from unbiased estimates can lead to biased estimators. For example, choosing from results of 50 randomized experiments is still not unbiased. Choosing based on "plausibility" is probably even worse. A 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 $\Rightarrow$ Model Dependence
The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion
The Problems Matching Solves

Without Matching:

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

Without Matching:

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

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

Without Matching:

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

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

Without Matching:

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

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

Without Matching:

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

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse[eff]
- conscientious effort doesn’t avoid biases (Banaji 2013)[acc]
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse
- Conscientious effort doesn’t avoid biases (Banaji 2013)
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)
The Problems Matching Solves

Without Matching:

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

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

Matching to Reduce Model Dependence
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 $\sim$ Model Dependence $\sim$ Researcher discretion $\sim$ 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 $\sim$ Model Dependence $\sim$ Researcher discretion $\sim$ Bias
The Problems Matching Solves

Without Matching:

- Imbalance ⊸ Model Dependence ⊸ Researcher discretion ⊸ Bias
The Problems Matching Solves

**Without Matching:**

- Imbalance $\sim$ Model Dependence $\sim$ Researcher discretion $\sim$ Bias

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

- **Notation:**
  - \( Y_i \): dependent variable
  - \( T_i \) (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?

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

- **Notation:** \( Y_i \) 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?

- **Notation**: \( Y_i \) 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?

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

\[
TE_i = Y_i(1) - Y_i(0) = \text{observed} - \text{unobserved}
\]
What’s Matching?

• **Notation:** $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders

• **Treatment Effect for treated observation $i$:**

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

= observed – unobserved
What’s Matching?

- **Notation:** $Y_i$ 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?

- **Notation:** $Y_i$ 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:

   $$SATT = \frac{1}{\sum_{i} I(T_i = 1)} \sum_{i: T_i = 1} (TE_i)$$

2. **FSATT:** Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias

Matching to Reduce Model Dependence
What’s Matching?

- **Notation**: $Y_i$ 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:

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

- **Notation**: $Y_i$ 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:
     \[
     \text{SATT} = \text{Mean} \left( \frac{TE_i}{i \in \{T_i=1\}} \right)
     \]
  2. **FSATT**: Feasible SATT (prune badly matched treateds too)
What’s Matching?

• **Notation:** $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders

• **Treatment Effect for treated observation $i$:**

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

= 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:

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

  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?

- **Notation:** $Y_i$ 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:
     \[
     \text{SATT} = \text{Mean} \left( \text{TE}_i \right)_{i \in \{T_i=1\}}
     \]
  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
Evaluating Reduction in Model Dependence

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time

- **Data:** $n = 408$ new drugs (262 approved, 146 pending)

- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)

- **Model:** lognormal survival

- **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)

- **Match:** prune 49 units (2 treated, 17 control units)

- **Run:** 262,143 possible specifications; calculate SATT for each

- **Evaluate:** Variability in SATT across specifications

(Normally we'd only use one or a few specifications)
Evaluating Reduction in Model Dependence

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- Data: $n = 408$ new drugs (262 approved, 146 pending)
- Measured confounders: 18 (clinical factors, firm characteristics, media variables, etc.)
- Model: lognormal survival
- QOI: Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
- Match: prune 49 units (2 treated, 17 control units)
- Run: 262,143 possible specifications; calculate SATT for each
- Evaluate: Variability in SATT across specifications

(Normally we'd only use one or a few specifications)
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis**: Democratic senate majorities slow FDA drug approval time
- **Data**: $n = 408$ new drugs (262 approved, 146 pending)
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis**: Democratic senate majorities slow FDA drug approval time
- **Data**: $n = 408$ new drugs (262 approved, 146 pending)
- **Measured confounders**: 18 (clinical factors, firm characteristics, media variables, etc.)
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
- **Data:** $n = 408$ new drugs (262 approved, 146 pending)
- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
- **Model:** lognormal survival
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
- **Data:** \( n = 408 \) new drugs (262 approved, 146 pending)
- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
- **Model:** lognormal survival
- **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
Evaluating Reduction in Model Dependence

Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
- **Data:** \( n = 408 \) new drugs (262 approved, 146 pending)
- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
- **Model:** lognormal survival
- **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
- **Match:** prune 49 units (2 treated, 17 control units)
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
- **Data:** $n = 408$ new drugs (262 approved, 146 pending)
- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
- **Model:** lognormal survival
- **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
- **Match:** prune 49 units (2 treated, 17 control units)
- **Run:** 262,143 possible specifications; calculate SATT for each
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

• **Hypothesis:** Democratic senate majorities slow FDA drug approval time
• **Data:** $n = 408$ new drugs (262 approved, 146 pending)
• **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
• **Model:** lognormal survival
• **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
• **Match:** prune 49 units (2 treated, 17 control units)
• **Run:** 262,143 possible specifications; calculate SATT for each
• **Evaluate:** *Variability* in SATT across specifications
Evaluating Reduction in Model Dependence
Empirical Illustration: Carpenter, AJPS, 2002

- **Hypothesis:** Democratic senate majorities slow FDA drug approval time
- **Data:** \( n = 408 \) new drugs (262 approved, 146 pending)
- **Measured confounders:** 18 (clinical factors, firm characteristics, media variables, etc.)
- **Model:** lognormal survival
- **QOI:** Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
- **Match:** prune 49 units (2 treated, 17 control units)
- **Run:** 262,143 possible specifications; calculate SATT for each
- **Evaluate:** *Variability* in SATT across specifications
  - (Normally we’d only use one or a few specifications)
Reducing Model Dependence

Point estimate of Carpenter's specification using raw data

Estimated in-sample average treatment effect for the treated

Density

0.00 0.05 0.10 0.15 0.20

Raw data

Matched data

Matching to Reduce Model Dependence
Reducing Model Dependence

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

![Graph](image-url)

- **Matched data**
- **Point estimate of raw data**
- **Raw data**

Matching to Reduce Model Dependence
Another Example: Jeffrey Koch, AJPS, 2002

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
- On average
- Exact
- 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
  - Unobserved
  - On average

Matching to Reduce Model Dependence

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

2. E.g., Imai, King, Nall 2009: SEs 600% smaller!
Types of Experiments

- Complete Randomization
- Fully Blocked
- Observed on average
- Unobserved on average

$\text{Fully blocked dominates complete randomization for:}
\begin{align*}
\text{imbalance,} & \\
\text{model dependence,} & \\
\text{power,} & \\
\text{efficiency,} & \\
\text{bias,} & \\
\text{research costs,} & \\
\text{robustness.} & \\
\end{align*}$

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

Complete Randomization

Matching to Reduce Model Dependence
Matching: Finding Hidden Randomized Experiments

Types of Experiments

| Complete Randomization | Fully Blocked |

跛化的随机化完全超越了完全随机化，例如，伊迈、金、纳尔 2009：误差减少了 600%！

根据观测数据的匹配目标

• PSM：完全随机化
• 其他方法：完全随机化
• 其他匹配方法优于 PSM
### Types of Experiments

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

- Full blocked dominates complete randomization for: imbalances, 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>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

E.g., Imai, King, Nall 2009: SEs 600% smaller!
### 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></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>
### 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></td>
</tr>
</tbody>
</table>
### 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>
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><strong>Observed</strong></td>
<td>On average</td>
<td>On average</td>
</tr>
<tr>
<td><strong>Unobserved</strong></td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

~ Fully blocked dominates complete randomization

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

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

\(\rightarrow\) **Fully blocked** dominates **complete randomization** 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</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>

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

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

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

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

~ **Fully blocked** dominates **complete randomization** for: imbalance, model dependence, power,
## Types of Experiments

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

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

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

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

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

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

#### Types of Experiments

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

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

#### Goal of Each Matching Method (in Observational Data)
## 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>(\text{Observed})</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>(\text{Unobserved})</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\(\nearrow\) 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</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>

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

\(\sim\) 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
### Types of Experiments

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

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

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

- PSM: complete randomization
- Other methods: fully blocked
- Other matching methods dominate PSM (wait, it gets worse)
Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier
Method 1: Mahalanobis Distance Matching

Procedure

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

Interpretation

- Quiz: Do you understand the distance trade offs?
- Quiz: Does standardization help?

Mahalanobis is for methodologists; in applications, use Euclidean!
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation
• Quiz: Do you understand the distance trade offs?
• Quiz: Does standardization help?
↝ Mahalanobis is for methodologists; in applications, use Euclidean!
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation
• Quiz: Do you understand the distance trade offs?
• Quiz: Does standardization help?
↝ Mahalanobis is for methodologists; in applications, use Euclidean!
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

1. **Preprocess** *(Matching)*

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

- 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)

Interpretation

- Quiz: Do you understand the distance trade offs?
- Quiz: Does standardization help?
  ↝ Mahalanobis is for methodologists; in applications, use Euclidean!
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

1. **Preprocess** *(Matching)*

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

Procedure

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)

Procedure

1. **Preprocess** (Matching)
   - Distance\((X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}\)
   - Match each treated unit to the nearest control unit

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

Procedure

1. **Preprocess** (Matching)
   - Distance \((X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}\)
   - 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 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

1. **Preprocess** (Matching)
   
   - Distance\((X_c, X_t) = \sqrt{(X_c - X_t)\prime S^{-1}(X_c - X_t)}\)
   - 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 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

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

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

Procedure

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

Interpretation

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

Procedure

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

Interpretation

- **Quiz:** Do you understand the distance trade-offs?
Method 1: Mahalanobis Distance Matching  
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation

- **Quiz:** Do you understand the distance trade offs?
- **Quiz:** Does standardization help?
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

Procedure

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

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

Interpretation

- **Quiz**: Do you understand the distance trade offs?
- **Quiz**: Does standardization help?

$\sim$ Mahalanobis is for methodologists; in applications, use Euclidean!
Mahalanobis Distance Matching

Education (years)

Age

Education (years)
Mahalanobis Distance Matching

Three Matching Methods
Three Matching Methods
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching

![Scatter plot showing relationship between Education (years) and Age.](chart.png)
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

**Interpretation**
- Quiz: Do you understand distance trade offs?
- Quiz: What do you do if you have too few observations?
Method 2: Coarsened Exact Matching
(Approximates FullyBlocked 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
   - Weight controls in each stratum to equal treateds

Interpretation
   - Quiz: Do you understand distance trade offs?
   - Quiz: What do you do if you have too few observations?
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation

- Quiz: Do you understand distance trade offs?
- Quiz: What do you do if you have too few observations?
## Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

### Procedure

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

### Interpretation

- Quiz: Do you understand distance trade offs?
- Quiz: What do you do if you have too few observations?
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

Procedure

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

Procedure

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)

Procedure

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)

Procedure

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)

Procedure

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)

Procedure

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

Interpretation

Quiz: Do you understand distance trade offs?
Quiz: What do you do if you have too few observations?
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation

- Quiz: Do you understand distance trade offs?
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

Procedure

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

Interpretation

- Quiz: Do you understand distance trade offs?
- Quiz: What do you do if you have too few observations?
Coarsened Exact Matching
### Coarsened Exact Matching

<table>
<thead>
<tr>
<th>Education</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing coarsened exact matching](image-url)

Three Matching Methods
Coarsened Exact Matching

Three Matching Methods
Coarsened Exact Matching

Three Matching Methods
Three Matching Methods
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching

Three Matching Methods
Best Case: Coarsened Exact Matching

Three Matching Methods
Method 3: Propensity Score Matching

Procedure

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

Interpretation

• Quiz: Do you understand distance trade offs?
• Quiz: What do you do when one variable is very important?
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

Interpretation
- Quiz: Do you understand distance trade offs?
- Quiz: What do you do when one variable is very important?
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

Procedure

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

Interpretation

• Quiz: Do you understand distance trade offs?
• Quiz: What do you do when one variable is very important?
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

Procedure

1. **Preprocess** *(Matching)*
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

Procedure

1. **Preprocess** *(Matching)*

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

Procedure

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

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

Procedure

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)

Procedure

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)

Procedure

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)

Procedure

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)

Procedure

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

Three Matching Methods
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

**Procedure**

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

**Interpretation**
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

Procedure

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

Interpretation

- **Quiz:** Do you understand distance trade offs?
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

Procedure

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

Interpretation

• Quiz: Do you understand distance trade offs?
• Quiz: What do you do when one variable is very important?
Propensity Score Matching

Three Matching Methods
Propensity Score Matching

Three Matching Methods
Propensity Score Matching

Three Matching Methods
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

Three Matching Methods
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching

Three Matching Methods
Best Case: Propensity Score Matching is Suboptimal
Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier
Random Pruning Increases Imbalance

• "Random pruning": pruning process is independent of $X$.

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

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

• Result is completely general.

Problems with Propensity Score Matching
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 \( \mapsto \) 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 \( \mapsto n \) declines \( \mapsto E(d^2) \) increases

  • \( \Rightarrow \) random pruning increases imbalance

• Result is completely general
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

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

Problems with Propensity Score Matching
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$
• Discrete 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 $M_c, F_c$
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

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

Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$
• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\leadsto$ 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
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

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

Deleting data only helps if you’re careful!

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

• **Discrete example**
  - Sex-balanced dataset: 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

• **Continuous example**
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\sim 4$ 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
- **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 $\sim$ 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

- **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 $\sim$ 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
- **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 $\sim n$ declines $\sim E(d^2)$ 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 $\sim$ 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

- 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
    $d = \overline{X}_t - \overline{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\sim n$ declines $\sim E(d^2)$ 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 $\leadsto$ 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 $\leadsto n$ declines $\leadsto E(d^2)$ increases
  - $\implies$ random pruning increases imbalance
- Result is completely general
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: $X_c = X_t \Rightarrow \pi_c = \pi_t \Rightarrow X_c = X_t$

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)
     $\Rightarrow \hat{\pi} \approx 0.5$ (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?
     Nope.

Problems with Propensity Score Matching
PSM’s Statistical Properties

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

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

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

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

2. **The PSM Paradox:** When you do “better,” you do worse
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
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:
     \[ X_c = X_t \Rightarrow \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

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

2. **The PSM Paradox: When you do “better,” you do worse**
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata)
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:
     \[ X_c = X_t \Rightarrow \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \not\Rightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random
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:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

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

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

2. **The PSM Paradox:** When you do “better,” you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) Imbalance \( \sim \) Inefficiency \( \sim \) 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:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

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

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

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

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

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

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) \( \Rightarrow \) all \( \hat{\pi} \approx 0.5 \) (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? Nope.
PSM’s Statistical Properties

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

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

Problems with Propensity Score Matching
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)

Problems with Propensity Score Matching
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
Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier
Tensions in Existing Matching Methods
Tensions in Existing Matching Methods

- Maximize one metric; judge against another: Propensity score matching, compared with var-by-var diff in means
Tensions in Existing Matching Methods

- **Maximize one metric; judge against another**: Propensity score matching, compared with var-by-var diff in means
- **Choose \( n \); check imbalance after**: Propensity score matching, Mahalanobis
Tensions in Existing Matching Methods

• Maximize one metric; judge against another: Propensity score matching, compared with var-by-var diff in means
• Choose $n$; check imbalance after: Propensity score matching, Mahalanobis
• Choose imbalance; check $n$ after: exact matching, CEM
A Solution: The Matching Frontier

Number of Units Pruned vs. Imbalance
A Solution: The Matching Frontier

Number of Units Pruned

Imbalance

Low variance

High variance

Less biased

More biased

Number of Units Pruned
A Solution: The Matching Frontier

![Graph showing the relationship between the number of units pruned and the variance, with a point labeled Result #1 at low variance and less biased.]

The Matching Frontier

40/44
A Solution: The Matching Frontier

Number of Units Pruned vs. Imbalance

- Low variance
  - Less biased
  - Result #1
  - Result #2
- High variance
  - More biased
A Solution: The Matching Frontier

The Matching Frontier
A Solution: The Matching Frontier

![Graph showing the relationship between number of units pruned and imbalance. The x-axis represents low variance to high variance, while the y-axis represents less biased to more biased. Four results are marked: Result #1, Result #2, Result #3, and Result #4.](image-url)
A Solution: The Matching Frontier

The graph shows the relationship between the number of units pruned and the imbalance in data. The x-axis represents the number of units pruned, with low variance on the left and high variance on the right. The y-axis represents the imbalance, with less biased on the bottom and more biased on the top.

- **Result #1**: Low variance, less biased
- **Result #2**: High variance, more biased
- **Result #3**: Low variance, less biased
- **Result #4**: High variance, more biased

The curved line illustrates the trade-off between variance and bias, indicating that as the number of units pruned increases, the variance also increases, leading to a more biased model.
A Solution: The Matching Frontier

- Theoretical frontier (optimal)
- Number of Units Pruned
- Imbalance
- Low variance High variance
- Less biased More biased
- Result #1
- Result #2
- Result #3
- Result #4
A Solution: The Matching Frontier

Number of Units Pruned
Imbalance

- Low variance
- High variance

- Less biased
- More biased

Result #1
Result #2
Result #3
Result #4

Theoretical frontier (optimal)

IMPOSSIBLE REGION
How hard is the frontier to calculate?

Consider 1 point on the SATT frontier:

Start with matrix of \( N \) control units \( X_0 \)

Calculate imbalance for all \((N \, n)\) subsets of rows of \( X_0 \)

Choose subset with lowest imbalance

Evaluations needed to compute the entire frontier:

\((N \, n)\) evaluations for each sample size \( n = N, N - 1, \ldots, 1 \)

The combination is the (gargantuan) "power set"

\( \Rightarrow \) 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:

  \( \text{Start with matrix of } N \text{ control units } X_0 \)

  \( \text{Calculate imbalance for all } (N^n) \text{ subsets of rows of } X_0 \)

  \( \text{Choose subset with lowest imbalance} \)

  \( \text{Evaluations needed to compute the entire frontier:} \)

  \( \text{Evaluations for each sample size } n = N, N-1, \ldots, 1 \)

  \( \text{The combination is the (gargantuan) “power set”} \)

  \( \text{e.g., } N > 300 \text{ requires more imbalance evaluations than } \)

  \( \text{elementary particles in the universe} \)

  \( \Rightarrow \text{It’s hard to calculate!} \)

  \( \text{We develop algorithms for the (optimal) frontier which:} \)

  \( \text{Runs very fast} \)

  \( \text{Operate as “greedy” but we prove are optimal} \)

  \( \text{Do not require evaluating every subset} \)

  \( \text{Work with very large data sets} \)

  \( \Rightarrow \text{It’s easy to calculate!} \)
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of \(N\) control units \(X_0\)
How hard is the frontier to calculate?

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

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

The Matching Frontier
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:
How hard is the frontier to calculate?

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

• Evaluations needed to compute the entire frontier:
  • \( \binom{N}{n} \) evaluations for each sample size \( n = N, N - 1, \ldots, 1 \)
How hard is the frontier to calculate?

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

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

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

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

Remaining Data

Covariate 1
Covariate 2

-1.0  -0.5  0.0  0.5  1.0

●

Treated
Control
Next to remove

Frontier

Number of Observations Dropped

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

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

The Matching Frontier
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

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

Number of Observations Dropped

Average Mahalanobis Discrepancy

-1.0 -0.5 0.0 0.5 1.0

-1.0 -0.5 0.0 0.5

-0.5 0.0 0.5

-0.5 0.0 0.5

Treated

Control

Next to remove

The Matching Frontier
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

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

The Matching Frontier
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

The Matching Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

○ Treated
□ Control
△ Next to remove

Covariate 1
Covariate 2
−1.0 −0.5 0.0 0.5 1.0

 Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped

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

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

The Matching Frontier
Constructing the FSATT Mahalanobis Frontier

### Remaining Data

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0</td>
<td>-0.5</td>
</tr>
<tr>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

- ○ Treated
- ○ Control
- □ Next to remove

### Frontier

- ● Treated
- ○ Control
- □ Next to remove

#### Number of Observations Dropped

- 0
- 5
- 10
- 15
- 20

#### Average Mahalanobis Discrepancy

- 0.0
- 0.1
- 0.2
- 0.3
- 0.4

#### Number of Observations Dropped

- 0
- 5
- 10
- 15
- 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- 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

0.0 0.1 0.2 0.3 0.4

0.0 0.1 0.2 0.3 0.4

Number of Observations Dropped

Average Mahalanobis Discrepancy

Treated
Control
Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

The Matching Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1

Covariate 2

-1.0, -0.5, 0.0, 0.5, 1.0

Treated
Control
Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped

0, 5, 10, 15, 20

0.0, 0.1, 0.2, 0.3, 0.4

The Matching Frontier
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

The Matching Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Warning: figure omits details and proof!

The Matching Frontier
Discrete algorithm

• Calculate bins
• Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

<table>
<thead>
<tr>
<th>Bin 1</th>
<th>Bin 2</th>
<th>Bin 3</th>
<th>Bin 4</th>
<th>Bin 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Discrete algorithm

Short version:
Discrete algorithm

Short version:

- Calculate bins
Discrete algorithm

Short version:

• Calculate bins
• Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
• Calculate bins
• Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

The Matching Frontier
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

![Graph showing bin distributions and number of observations pruned]
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

![Graph showing bins and observations pruned]
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

![Bar chart showing bins][1]

![Line chart showing number of observations pruned][2]
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

The Matching Frontier
Discrete algorithm

Short version:
• Calculate bins
• Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Discrete algorithm

Short version:

- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
• 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