Simplifying Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

(Talk at the U.S. Food and Drug Administration [CBER/OBE], 8/25/2015)

1Based on joint work with Rich Nielsen, Chris Lucas, Stefano Iacus, and Giuseppe Porro

2GaryKing.org
Imbalance $\Rightarrow$ Model Dependence $\Rightarrow$ Researcher Discretion $\Rightarrow$ Bias
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?
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
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?
Two Logit Models, Apparently Similar Results

<table>
<thead>
<tr>
<th></th>
<th>Original &quot;Interactive&quot; Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Coeff</td>
<td>SE</td>
</tr>
<tr>
<td>Wartype</td>
<td>−1.742</td>
<td>.609</td>
</tr>
<tr>
<td>Logdead</td>
<td>−.445</td>
<td>.126</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>Factnum</td>
<td>−1.259</td>
<td>.703</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.062</td>
<td>.065</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>Exp</td>
<td>−6.016</td>
<td>.071</td>
</tr>
<tr>
<td>Decade</td>
<td>−.299</td>
<td>.169</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td></td>
<td>.037</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 R²</td>
<td>.423</td>
<td></td>
</tr>
</tbody>
</table>
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

<table>
<thead>
<tr>
<th></th>
<th>Original &quot;Interactive&quot; Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wartype</td>
<td>$-1.742$, $p = 0.004$</td>
<td>$-1.666$, $p = 0.006$</td>
</tr>
<tr>
<td>Logdead</td>
<td>$-0.445$, $p = 0.000$</td>
<td>$-0.437$, $p = 0.000$</td>
</tr>
<tr>
<td>Wardur</td>
<td>$0.006$, $p = 0.258$</td>
<td>$0.006$, $p = 0.006$</td>
</tr>
<tr>
<td>Factnum</td>
<td>$-1.259$, $p = 0.073$</td>
<td>$-1.045$, $p = 0.245$</td>
</tr>
<tr>
<td>Factnum2</td>
<td>$0.062$, $p = 0.346$</td>
<td>$0.065$, $p = 0.104$</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>$0.004$, $p = 0.010$</td>
<td>$0.004$, $p = 0.017$</td>
</tr>
<tr>
<td>Develop</td>
<td>$0.001$, $p = 0.065$</td>
<td>$0.001$, $p = 0.068$</td>
</tr>
<tr>
<td>Exp</td>
<td>$-6.016$, $p = 0.050$</td>
<td>$-6.215$, $p = 0.043$</td>
</tr>
<tr>
<td>Decade</td>
<td>$-0.299$, $p = 0.077$</td>
<td>$-0.284$, $p = 0.093$</td>
</tr>
<tr>
<td>Treaty</td>
<td>$2.124$, $p = 0.010$</td>
<td>$2.126$, $p = 0.008$</td>
</tr>
<tr>
<td>UNOP4</td>
<td>$3.135$, $p = 0.004$</td>
<td>$2.262$, $p = 0.851$</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td></td>
<td>$-0.037$, $p = 0.011$</td>
</tr>
<tr>
<td>Constant</td>
<td>$8.609$, $p = 0.000$</td>
<td>$7.978$, $p = 0.000$</td>
</tr>
</tbody>
</table>

N = 122

Log-likelihood: -45.649, -44.902

Pseudo $R^2$: 0.423, 0.433
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

<table>
<thead>
<tr>
<th>Variables</th>
<th>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>
Model Dependence: Same Fit, Different Predictions

In Sample Fit

Counterfactual Prediction

![Graph showing model dependence with same fit, different predictions.](image-url)
Coarsened Exact Matching
Coarsened Exact Matching (CEM)

1. Preprocess \((X, T)\) with CEM:
   - (A) Temporarily coarsen \(X\) as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as a histogram
   - (B) Perform exact matching on 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
   - (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   - Or apply other matching methods within CEM strata & they inherit CEM's properties

⇾ A version of CEM: Last studied 45 years ago by Cochran
⇾ First used many decades before that
⇾ We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   
   (A) Temporarily coarsen \(X\) as much as you’re willing
   - e.g., Education (grade school, high school, college, graduate)
   - Easy to understand, or can be automated as a histogram
   
   (B) Perform exact matching on 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
   
   (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   - Or apply other matching methods within CEM strata & they inherit CEM’s properties

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
\(\Rightarrow\) We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:

   (A) Temporarily coarsen \(X\) as much as you're willing
      - e.g., Education (grade school, high school, college, graduate)
      - Easy to understand, or can be automated as a histogram

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   - Or apply other matching methods within CEM strata & they inherit CEM's properties

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
\(\Rightarrow\) We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess $(X, T)$ with CEM:
   (A) Temporarily coarsen $X$ as much as you’re willing

The method is:

- Easy to understand, or can be automated as a histogram
- Simple to implement
- Prone to missing data issues
- Robust to model misspecification
- Predicts causal effects, not just associations
- Performs well in practice

The method can be extended in various ways, and it is a useful tool for causal inference.
Coarsened Exact Matching (CEM)
A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
      • e.g., Education (grade school, high school, college, graduate)

2. Analyze as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherit CEM’s properties)

⇝ A version of CEM: Last studied 45 years ago by Cochran
⇝ First used many decades before that
⇝ We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. **Preprocess** \((X, T)\) with CEM:
   
   (A) Temporarily coarsen \(X\) as much as you’re willing
   
   - e.g., Education (grade school, high school, college, graduate)
   - Easy to understand, or can be automated as a histogram

2. Analyze as without matching (adding weights for stratum-size)
   
   - Or apply other matching methods within CEM strata & they inherit CEM’s properties

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran

\(\Rightarrow\) First used many decades before that

\(\Rightarrow\) We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)
A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you're willing
       - e.g., Education (grade school, high school, college, graduate)
       - Easy to understand, or can be automated as a histogram
   (B) Perform exact matching on the coarsened \(X, C(X)\)
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
       - e.g., Education (grade school, high school, college, graduate)
       - Easy to understand, or can be automated as a histogram
   (B) Perform exact matching on the coarsened \(X, C(X)\)
       - Sort observations into strata, each with unique values of \(C(X)\)

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
\(\Rightarrow\) We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. **Preprocess** \((X, T)\) with CEM:
   
   (A) **Temporarily coarsen** \(X\) as much as you’re willing
   
   - e.g., Education (grade school, high school, college, graduate)
   - Easy to understand, or can be automated as a histogram

   (B) **Perform exact matching** on 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
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. **Preprocess** \((X, T)\) with CEM:
   
   (A) Temporarily coarsen \(X\) as much as you’re willing
   
   - e.g., Education (grade school, high school, college, graduate)
   - Easy to understand, or can be automated as a histogram

   (B) Perform **exact matching** on 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

   (C) Pass on original (uncoarsened) units except those pruned
Coarsened Exact Matching (CEM)
A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
       • e.g., Education (grade school, high school, college, graduate)
       • Easy to understand, or can be automated as a histogram
   (B) Perform exact matching on 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
   (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   - (A) Temporarily coarsen \(X\) as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as a histogram
   - (B) Perform **exact matching** on 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
   - (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherert CEM’s properties)
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
      • e.g., Education (grade school, high school, college, graduate)
      • Easy to understand, or can be automated as a histogram
   (B) Perform exact matching on 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
   (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherit CEM’s properties)

\(\rightsquigarrow\) A version of CEM: Last studied 45 years ago by Cochran
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
   * e.g., Education (grade school, high school, college, graduate)
   * Easy to understand, or can be automated as a histogram
   (B) Perform exact matching on 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
   (C) Pass on original (uncoarsened) units except those pruned

2. Analyze as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherert CEM’s properties)

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. **Preprocess** \((X, T)\) with CEM:
   (A) Temporarily coarsen \(X\) as much as you’re willing
   - e.g., Education (grade school, high school, college, graduate)
   - Easy to understand, or can be automated as a histogram
   (B) Perform **exact matching** on 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
   (C) Pass on original (uncoarsened) units except those pruned

2. **Analyze** as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherit CEM’s properties)

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
\(\Rightarrow\) We prove: many new properties, uses, & extensions,
Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. **Preprocess** \((X, T)\) with CEM:
   - **(A) Temporarily coarsen** \(X\) as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as a histogram
   - **(B) Perform exact matching** on 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
   - **(C) Pass on original (uncoarsened) units** except those pruned

2. **Analyze** as without matching (adding weights for stratum-size)
   (Or apply other matching methods within CEM strata & they inherent CEM’s properties)

\(\Rightarrow\) A version of CEM: Last studied 45 years ago by Cochran
\(\Rightarrow\) First used many decades before that
\(\Rightarrow\) We prove: many new properties, uses, & extensions, and show how it resolves many problems in the literature
Problems with Other Matching Methods

• Don't eliminate the extrapolation region
• Don't work with multiply imputed data
• Violate the congruence principle
• Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data:
  - require normal data (or DMPES);
  - all $X$'s must have same effect on $Y$;
  - $Y$ must be a linear function of $X$;
  - aim only for expected (not in-sample) imbalance;
  - $\Rightarrow$ in practice, we're lucky if mean imbalance is reduced
• Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
• Hard to use: Improving balance on 1 variable can reduce it on others
• Best practice:
  - choose $n$-
  - match-
  - check,
  - tweak-
  - match-
  - check,
  - tweak-
  - match-
  - check,
  - ···
• Actual practice:
  - choose $n$,
  - match,
  - publish,
  - STOP.
  - (Is balance even improved?)
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data

Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data:

- require normal data (or DMPES);
- all $X$’s must have same effect on $Y$;
- $Y$ must be a linear function of $X$;
- aim only for expected (not in-sample) imbalance;
- $\Rightarrow$ in practice, we’re lucky if mean imbalance is reduced

Not well designed for observational data:

- Least important (variance): matched $n$ chosen ex ante
- Most important (bias): imbalance reduction checked ex post

- Hard to use: Improving balance on 1 variable can reduce it on others

Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweaking-match-check, ···

Actual practice: choose $n$, match, publish, STOP. (Is balance even improved?)
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data:
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES);
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$;

...
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$;
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance;
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\leadsto$ in practice, we’re lucky if mean imbalance is reduced
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; \( \rightsquigarrow \) in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; \(\rightsquigarrow\) in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \(n\) chosen \textit{ex ante}
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \( n \) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
Problems with Other Matching Methods

• Don’t eliminate the extrapolation region
• Don’t work with multiply imputed data
• Violate the congruence principle
• Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$‘s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
• Not well designed for observational data:
  • Least important (variance): matched $n$ chosen ex ante
  • Most important (bias): imbalance reduction checked ex post
• Hard to use: Improving balance on 1 variable can reduce it on others
  • Best practice:
Problems with Other Matching Methods

• Don’t eliminate the extrapolation region
• Don’t work with multiply imputed data
• Violate the congruence principle
• Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
• Not well designed for observational data:
  • Least important (variance): matched $n$ chosen ex ante
  • Most important (bias): imbalance reduction checked ex post
• Hard to use: Improving balance on 1 variable can reduce it on others
  • Best practice: choose $n$
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; ↝ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check,
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; ⇾ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \( n \) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose \( n \)-match-check, tweak-match-check,
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\sim$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \( n \) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check,
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all \(X\)'s must have same effect on \(Y\); \(Y\) must be a linear function of \(X\); aim only for expected (not in-sample) imbalance; \(\rightsquigarrow\) in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \(n\) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose \(n\)-match-check, tweak-match-check, tweak-match-check, tweak
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; \( \leadsto \) in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \( n \) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose \( n \)-match-check, tweak-match-check, tweak-match-check, tweak-match
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; ↝ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check,
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, …
Problems with Other Matching Methods

• Don’t eliminate the extrapolation region
• Don’t work with multiply imputed data
• Violate the congruence principle
• Matching methods from the largest class (EPBR, e.g., PSM, MDM) *don’t apply to real data*: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\leadsto$ in practice, we’re lucky if mean imbalance is reduced
• Not well designed for observational data:
  • Least important (variance): matched $n$ chosen *ex ante*
  • Most important (bias): imbalance reduction checked *ex post*
• Hard to use: Improving balance on 1 variable can reduce it on others
  • Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, · · ·
  • Actual practice:
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; \( \Rightarrow \) in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched \( n \) chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose \( n \)-match-check, tweak-match-check, tweak-match-check, \( \cdots \)
  - Actual practice: choose \( n \),
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\leadsto$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, ···
  - Actual practice: choose $n$, match,
Problems with Other Matching Methods

• Don’t eliminate the extrapolation region
• Don’t work with multiply imputed data
• Violate the congruence principle
• Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all X’s must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; ⇝ in practice, we’re lucky if mean imbalance is reduced
• Not well designed for observational data:
  • Least important (variance): matched \( n \) chosen ex ante
  • Most important (bias): imbalance reduction checked ex post
• Hard to use: Improving balance on 1 variable can reduce it on others
  • Best practice: choose \( n \)-match-check, tweak-match-check, tweak-match-check, tweak-match-check, · · ·
  • Actual practice: choose \( n \), match, publish,
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$, match, publish, STOP.
Problems with Other Matching Methods

- Don’t eliminate the extrapolation region
- Don’t work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don’t apply to real data: require normal data (or DMPES); all $X$’s must have same effect on $Y$; $Y$ must be a linear function of $X$; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we’re lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, · · ·
  - Actual practice: choose $n$, match, publish, STOP. (Is balance even improved?)
What’s Coarsening?

- Coarsening is intrinsic to measurement.
- We think of measurement as clarity between categories.
- But measurement also involves homogeneity within categories.
- Examples: male/female, rich/middle/poor, black/white, war/nonwar.

Better measurements (e.g., telescopes) $\Rightarrow$ better resolution.

Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:

- 7 point Party ID $\Rightarrow$ Democrat/Independent/Republican
- Likert Issue questions $\Rightarrow$ agree/neutral/no opinion/disagree
- Multiparty voting $\Rightarrow$ winner/losers
- Religion, Occupation, SEC industries, ICD codes, etc.

Temporary Coarsening for CEM; e.g.:

- Education: grade school, middle school, high school, college, graduate
- Income: poverty level threshold, or larger bins for higher income
- Age: infant, child, adolescent, young adult, middle age, elderly
What’s Coarsening?

- Coarsening is intrinsic to measurement
What’s Coarsening?

• Coarsening is intrinsic to measurement
  • We think of measurement as clarity between categories
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as *clarity between categories*
  - But measurement also involves *homogeneity within categories*

Examples:
- Male/female, rich/middle/poor, black/white, war/nonwar.
- Better measurements (e.g., telescopes) \(\rightarrow\) better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID \(\rightarrow\) Democrat/Independent/Republican
  - Likert Issue questions \(\rightarrow\) agree/neutral/no opinion/disagree
  - Multiparty voting \(\rightarrow\) winner/losers
- Religion, Occupation, SEC industries, ICD codes, etc.

Temporary Coarsening for CEM; e.g.:
- Education: grade school, middle school, high school, college, graduate
- Income: poverty level threshold, or larger bins for higher income
- Age: infant, child, adolescent, young adult, middle age, elderly
What’s Coarsening?

• Coarsening is intrinsic to measurement
  • We think of measurement as clarity between categories
  • But measurement also involves homogeneity within categories
  • Examples: male/female, rich/middle/poor, black/white, war/nonwar.
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\leadsto$ better resolution
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\sim$ better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
What’s Coarsening?

- **Coarsening is intrinsic to measurement**
  - We think of measurement as **clarity between categories**
  - But measurement also involves **homogeneity within categories**
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) ⇒ better resolution

- **Data analysts routinely coarsen**, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID ⇒ Democrat/Independent/Republican
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) \(\leadsto\) better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID \(\leadsto\) Democrat/Independent/Republican
  - Likert Issue questions \(\leadsto\) agree/\{neutral,no opinion\}/disagree
What’s Coarsening?

- **Coarsening is intrinsic to measurement**
  - We think of measurement as *clarity between categories*
  - But measurement also involves *homogeneity within categories*
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\Rightarrow$ better resolution

- **Data analysts routinely coarsen**, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\Rightarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\Rightarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\Rightarrow$ winner/losers
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\Rightarrow$ better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\Rightarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\Rightarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\Rightarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) \(\Rightarrow\) better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID \(\rightsquigarrow\) Democrat/Independent/Republican
  - Likert Issue questions \(\rightsquigarrow\) agree/\{neutral, no opinion\}/disagree
  - multiparty voting \(\rightsquigarrow\) winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) \(\mapsto\) better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID \(\mapsto\) Democrat/Independent/Republican
  - Likert Issue questions \(\mapsto\) agree/\{neutral,no opinion\}/disagree
  - multiparty voting \(\mapsto\) winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
  - Education: grade school, middle school, high school, college, graduate
What’s Coarsening?

- **Coarsening is intrinsic to measurement**
  - We think of measurement as **clarity between categories**
  - But measurement also involves **homogeneity within categories**
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\leadsto$ better resolution

- **Data analysts routinely coarsen**, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\leadsto$ Democrat/Independent/Republican
  - Likert Issue questions $\leadsto$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\leadsto$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.

- **Temporary Coarsening for CEM**; e.g.:
  - Education: grade school, middle school, high school, college, graduate
  - Income: poverty level threshold, or larger bins for higher income
What’s Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurements (e.g., telescopes) $\Rightarrow$ better resolution
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\Rightarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\Rightarrow$ agree/\{neutral,no opinion\}/disagree
  - multiparty voting $\Rightarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
  - Education: grade school, middle school, high school, college, graduate
  - Income: poverty level threshold, or larger bins for higher income
  - Age: infant, child, adolescent, young adult, middle age, elderly
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for:
  - means,
  - variances,
  - skewness,
  - covariances,
  - comoments,
  - coskewness,
  - co-kurtosis,
  - quantiles,
  - and full multivariate histogram.
- Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set...
  - too large? $\Rightarrow$ You’re left modeling remaining imbalances
  - too small? $\Rightarrow$ $n$ may be too small
  - as large as you’re comfortable with, but $n$ is still too small? $\Rightarrow$ No magic method of matching can save you; $\Rightarrow$ You’re stuck modeling or collecting better data
CEM as a Monotonic Imbalance Bounding Method

• Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)

- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for:
  - means,
  - variances,
  - skewness,
  - covariances,
  - comoments,
  - coskewness,
  - co-kurtosis,
  - quantiles,
  - and full multivariate histogram.

$\Rightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)

- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence

• What if $\epsilon$ is set . . .
  - too large? $\Rightarrow$ You're left modeling remaining imbalances
  - too small? $\Rightarrow n$ may be too small
  - as large as you're comfortable with, but $n$ is still too small? $\Rightarrow$ No magic method of matching can save you; $\Rightarrow$ You're stuck modeling or collecting better data
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for:
  - means,
  - variances,
  - skewness,
  - covariances,
  - comoments,
  - coskewness,
  - co-kurtosis,
  - quantiles,
  - and full multivariate histogram.

- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large? $\Rightarrow$ You're left modeling remaining imbalances
  - too small? $\Rightarrow n$ may be too small
  - as large as you're comfortable with, but $n$ is still too small? $\Rightarrow$ No magic method of matching can save you; $\Rightarrow$ You're stuck modeling or collecting better data
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means,
CEM as a Monotonic Imbalance Bounding Method

- Define: \( \epsilon \) as largest (coarsened) bin size (\( \epsilon = 0 \) is exact matching)
- We Prove: setting \( \epsilon \) bounds the treated-control group difference, within strata and globally, for: means, variances,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances,
CEM as a Monotonic Imbalance Bounding Method

• Define: \( \epsilon \) as largest (coarsened) bin size (\( \epsilon = 0 \) is exact matching)
• We Prove: setting \( \epsilon \) bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles,
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

• By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
• What if $\epsilon$ is set . . .
  - too large? $\Rightarrow$ You’re left modeling remaining imbalances
  - too small? $\Rightarrow$ $n$ may be too small
  - as large as you’re comfortable with, but $n$ is still too small? $\Rightarrow$ No magic method of matching can save you; $\Rightarrow$ You’re stuck modeling or collecting better data

11/25
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

$\Rightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Rightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\implies$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\implies$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large?
  - too small? $n$ may be too small
  - as large as you’re comfortable with, but $n$ is still too small?
  - No magic method of matching can save you; $\implies$ You’re stuck modeling or collecting better data
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\implies$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large? $\rightsquigarrow$ You’re left modeling remaining imbalances
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  \[ \Rightarrow \text{Setting } \epsilon \text{ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched } n \text{ is determined ex post)} \]
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large? $\Rightarrow$ You’re left modeling remaining imbalances
  - too small?

11/25
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  \[ \implies \text{Setting } \epsilon \text{ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched } n \text{ is determined } \text{ex post)} \]
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large? $\rightsquigarrow$ You’re left modeling remaining imbalances
  - too small? $\rightsquigarrow n$ may be too small
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)

- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Rightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)

- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence

- What if $\epsilon$ is set . . .
  - too large? $\rightsquigarrow$ You’re left modeling remaining imbalances
  - too small? $\rightsquigarrow$ $n$ may be too small
  - as large as you’re comfortable with, but $n$ is still too small?
CEM as a Monotonic Imbalance Bounding Method

- Define: $\epsilon$ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  \[ \Rightarrow \text{ Setting } \epsilon \text{ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched } n \text{ is determined ex post) } \]
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if $\epsilon$ is set . . .
  - too large? $\leadsto$ You’re left modeling remaining imbalances
  - too small? $\leadsto$ $n$ may be too small
  - as large as you’re comfortable with, but $n$ is still too small? $\leadsto$ No magic method of matching can save you;
CEM as a Monotonic Imbalance Bounding Method

- Define: \( \epsilon \) as largest (coarsened) bin size (\( \epsilon = 0 \) is exact matching)
- We Prove: setting \( \epsilon \) bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  \( \Rightarrow \) Setting \( \epsilon \) controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched \( n \) is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if \( \epsilon \) is set . . .
  - too large? \( \sim \) You’re left modeling remaining imbalances
  - too small? \( \sim \) \( n \) may be too small
  - as large as you’re comfortable with, but \( n \) is still too small?
    \( \sim \) No magic method of matching can save you;
    \( \sim \) You’re stuck modeling or collecting better data
Other CEM properties we prove

• Automatically eliminates extrapolation region (no separate step)
• Bounds: model dependence, researcher discretion, bias, estimation error
• Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
• CEM: \( \epsilon_j \) is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
• Approximate invariance to measurement error: CEM PSM MDM Genetic
  - Common Units 96.5 70.2 80.9 80.0
• Fast and memory-efficient even for large \( n \); can be fully automated
• Simple to teach: coarsen, then exact match
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)

• Bounds: model dependence, researcher discretion, bias, estimation error

• Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive

• CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett

• Approximate invariance to measurement error: CEM PSM MDM Genetic
  - % Common Units 96.5 70.2 80.9 80.0

• Fast and memory-efficient even for large $n$; can be fully automated

• Simple to teach: coarsen, then exact match
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds:** model dependence, researcher discretion, bias, estimation error
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds:** model dependence, researcher discretion, bias, estimation error
- Meets the *congruence principle*

---

Approximate invariance to measurement error: CEM PSM MDM Genetic
- % Common Units 96.5 70.2 80.9 80.0

- Fast and memory-efficient even for large \( n \); can be fully automated
- Simple to teach: coarsen, then exact match
Other CEM properties we prove

- **Automatically eliminates extrapolation region** (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- **Meets the congruence principle**
  - The principle: data space = analysis space

Approximate invariance to measurement error:

- CEM
- PSM
- MDM
- Genetic

<table>
<thead>
<tr>
<th>Common Units</th>
<th>96.5</th>
<th>70.2</th>
<th>80.9</th>
<th>80.0</th>
</tr>
</thead>
</table>

Fast and memory-efficient even for large $n$; can be fully automated

Simple to teach: coarsen, then exact match
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- Meets the *congruence principle*
  - The principle: data space $=$ analysis space
  - Estimators that violate it are nonrobust and counterintuitive
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds:** model dependence, researcher discretion, bias, estimation error
- **Meets the congruence principle**
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - **CEM:** $\epsilon_j$ is set using each variable’s units

Approximate invariance to measurement error:

<table>
<thead>
<tr>
<th>Common Units</th>
<th>96.5</th>
<th>70.2</th>
<th>80.9</th>
<th>80.0</th>
</tr>
</thead>
</table>

Fast and memory-efficient even for large $n$; can be fully automated

Simple to teach: coarsen, then exact match
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds:** model dependence, researcher discretion, bias, estimation error
- Meets the **congruence principle**
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit):

<table>
<thead>
<tr>
<th>Common Units</th>
<th>96.5</th>
<th>70.2</th>
<th>80.9</th>
<th>80.0</th>
</tr>
</thead>
</table>

- Fast and memory-efficient even for large $n$; can be fully automated
- Simple to teach: coarsen, then exact match
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student;
Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- **Bounds:** model dependence, researcher discretion, bias, estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett

Approximate invariance to measurement error: CEM PSM MDM Genetic

| Common Units | 96.5 | 70.2 | 80.9 | 80.0 |

Fast and memory-efficient even for large $n$; can be fully automated

Simple to teach: coarsen, then exact match
Other CEM properties we prove

- **Automatically eliminates extrapolation region** (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett

- **Approximate invariance to measurement error**:

<table>
<thead>
<tr>
<th>Method</th>
<th>CEM</th>
<th>PSM</th>
<th>MDM</th>
<th>Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Common Units</td>
<td>96.5</td>
<td>70.2</td>
<td>80.9</td>
<td>80.0</td>
</tr>
</tbody>
</table>
Other CEM properties we prove

- **Automatically eliminates extrapolation region** (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- **Meets the congruence principle**
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- **Approximate invariance to measurement error**:  
<table>
<thead>
<tr>
<th></th>
<th>CEM</th>
<th>PSM</th>
<th>MDM</th>
<th>Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Common Units</td>
<td>96.5</td>
<td>70.2</td>
<td>80.9</td>
<td>80.0</td>
</tr>
</tbody>
</table>
- **Fast and memory-efficient** even for large $n$; can be fully automated
Other CEM properties we prove

- **Automatically eliminates extrapolation region** (no separate step)
- **Bounds**: model dependence, researcher discretion, bias, estimation error
- **Meets the congruence principle**
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable’s units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- **Approximate invariance to measurement error**:
  - CEM  PSM  MDM  Genetic
  - % Common Units  96.5  70.2  80.9  80.0
- **Fast and memory-efficient even for large $n$; can be fully automated**
- **Simple to teach**: coarsen, then exact match
Monte Carlo: \( X \sim N(5, \Sigma) \) and \( X_C \sim N(5, \Sigma) \).

\( n = 2,000 \), \( \text{reps} = 5,000 \); Allow MDM & PSM to match with replacement; use automated CEM.

<table>
<thead>
<tr>
<th>Difference in means</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDM</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>PSM</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>CEM</td>
<td>0.04</td>
<td>0.04</td>
<td>0.08</td>
<td>0.06</td>
<td>0.07</td>
</tr>
</tbody>
</table>

\( \Rightarrow \) CEM dominates EPBR-methods in EPBR Data.
Monte Carlo:

$X \sim N(0, \Sigma)$ and $X \sim N(1, \Sigma)$. 

$n = 2,000, \text{ reps}=5,000$; allow MDM & PSM to match with replacement; use automated CEM.

<table>
<thead>
<tr>
<th>Difference in means</th>
<th>X_1</th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
<th>X_5</th>
<th>L_1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
</tr>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
</tr>
<tr>
<td>PSM</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.31</td>
</tr>
<tr>
<td>CEM</td>
<td>.04</td>
<td>.04</td>
<td>.08</td>
<td>.06</td>
<td>.07</td>
<td>.21</td>
</tr>
</tbody>
</table>

$\Rightarrow$ CEM dominates EPBR-methods in EPBR Data.
CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. 
Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000;
Monte Carlo: $\mathbf{X}_T \sim \mathcal{N}_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim \mathcal{N}_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement;
Monte Carlo: $X_T \sim N_5(\mathbf{0}, \Sigma)$ and $X_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps = 5,000; Allow MDM & PSM to match with replacement; use automated CEM
CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim \mathcal{N}_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim \mathcal{N}_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

Difference in means
CEM in Practice: EPBR-Compliant Data

Monte Carlo: $X_T \sim \mathcal{N}_5(\mathbf{0}, \Sigma)$ and $X_C \sim \mathcal{N}_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

Difference in means
Monte Carlo: $X_T \sim N_5(0, \Sigma)$ and $X_C \sim N_5(1, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th>Difference in means</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$L_1$</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
<td>.28</td>
</tr>
<tr>
<td>PSM</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.31</td>
<td>.16</td>
</tr>
<tr>
<td>CEM</td>
<td>.04</td>
<td>.04</td>
<td>.08</td>
<td>.06</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
</tr>
</tbody>
</table>
CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$L_1$</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
</tr>
</tbody>
</table>

$\Rightarrow$ CEM dominates EPBR-methods in EPBR Data
Monte Carlo: \( \mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma) \) and \( \mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma) \). \( n = 2,000, \ rep = 5,000; \) Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th></th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
<th>( L_1 )</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
<td>.28</td>
</tr>
</tbody>
</table>

CEM in Practice: EPBR-Compliant Data
CEM in Practice: EPBR-Compliant Data

Monte Carlo: \( \mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma) \) and \( \mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma) \). \( n = 2,000, \) reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th></th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
<th>( L_1 )</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
<td>.28</td>
</tr>
<tr>
<td>PSM</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.31</td>
<td>.16</td>
</tr>
</tbody>
</table>

\( \Rightarrow \) CEM dominates EPBR-methods in EPBR Data
Monte Carlo: \( \mathbf{X}_T \sim \mathcal{N}_5(\mathbf{0}, \Sigma) \) and \( \mathbf{X}_C \sim \mathcal{N}_5(\mathbf{1}, \Sigma) \). \( n = 2,000 \), \( \text{reps}=5,000 \); Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th>Difference in means</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
<th>( L_1 )</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
<td>.28</td>
</tr>
<tr>
<td>PSM</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.31</td>
<td>.16</td>
</tr>
<tr>
<td>CEM</td>
<td>.04</td>
<td>.04</td>
<td>.08</td>
<td>.06</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
</tr>
</tbody>
</table>
Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000; Allow MDM & PSM to match with replacement; use automated CEM

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$L_1$</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>MDM</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.34</td>
<td>.28</td>
</tr>
<tr>
<td>PSM</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.31</td>
<td>.16</td>
</tr>
<tr>
<td>CEM</td>
<td>.04</td>
<td>.04</td>
<td>.08</td>
<td>.06</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
</tr>
</tbody>
</table>

$\Rightarrow$ CEM dominates EPBR-methods in EPBR Data
CEM in Practice: Non-EPBR Data


<table>
<thead>
<tr>
<th>Method</th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00 1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03 1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02 1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38 1.12</td>
</tr>
<tr>
<td>CEM</td>
<td>.8</td>
<td>111.4</td>
<td>111.4</td>
<td>.03  .76</td>
</tr>
</tbody>
</table>

⇝

CEM works well in non-EPBR data too.
CEM in Practice: Non-EPBR Data

**Monte Carlo:** Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

<table>
<thead>
<tr>
<th></th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>−423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>02</td>
<td>1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38</td>
<td>1.12</td>
</tr>
<tr>
<td>CEM</td>
<td>.8</td>
<td>111.4</td>
<td>111.4</td>
<td>03</td>
<td>.76</td>
</tr>
</tbody>
</table>

CEM works well in non-EPBR data too.
**CEM in Practice: Non-EPBR Data**


<table>
<thead>
<tr>
<th>Method</th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02</td>
<td>1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38</td>
<td>1.12</td>
</tr>
<tr>
<td>CEM</td>
<td>.8</td>
<td>111.4</td>
<td>111.4</td>
<td>.03</td>
<td>.76</td>
</tr>
</tbody>
</table>

CEM works well in non-EPBR data too.
**Monte Carlo:** Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

<table>
<thead>
<tr>
<th></th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>−423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
</tbody>
</table>
CEM in Practice: Non-EPBR Data


<table>
<thead>
<tr>
<th></th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>$-423.7$</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>-423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02</td>
<td>1.23</td>
</tr>
</tbody>
</table>

CEM works well in non-EPBR data too.

<table>
<thead>
<tr>
<th>Method</th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>-423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02</td>
<td>1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38</td>
<td>1.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>$-423.7$</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02</td>
<td>1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38</td>
<td>1.12</td>
</tr>
<tr>
<td>CEM</td>
<td>.8</td>
<td>111.4</td>
<td>111.4</td>
<td>.03</td>
<td>.76</td>
</tr>
</tbody>
</table>
CEM in Practice: Non-EPBR Data


<table>
<thead>
<tr>
<th></th>
<th>BIAS</th>
<th>SD</th>
<th>RMSE</th>
<th>Seconds</th>
<th>$L_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>−423.7</td>
<td>1566.5</td>
<td>1622.6</td>
<td>.00</td>
<td>1.28</td>
</tr>
<tr>
<td>MDM</td>
<td>784.8</td>
<td>737.9</td>
<td>1077.2</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>PSM</td>
<td>260.5</td>
<td>1025.8</td>
<td>1058.4</td>
<td>.02</td>
<td>1.23</td>
</tr>
<tr>
<td>GEN</td>
<td>78.3</td>
<td>499.5</td>
<td>505.6</td>
<td>27.38</td>
<td>1.12</td>
</tr>
<tr>
<td>CEM</td>
<td>.8</td>
<td>111.4</td>
<td>111.4</td>
<td>.03</td>
<td>.76</td>
</tr>
</tbody>
</table>

→ CEM works well in non-EPBR data too
CEM Extensions I

- Many binary variables: coarsen sums of related variables.
- Missing Data and/or Measurement Error:
  1. Multiply Impute (missing) or Overimpute (measurement error).
  2. Put missing observation in stratum where plurality of imputations fall.
  3. Pass on uncoarsened imputations to analysis stage.
  4. Use the usual MI combining rules to analyze.
- Multicategory treatments: No modification necessary; keep all strata with \( \geq 1 \) unit having each value of \( T \).
- Blocking in Randomized Experiments: no modification needed; randomly assign \( T \) within CEM strata.
- Automating user choices:
  - Histogram bin size calculations,
  - Estimated SATT error bound,
  - Progressive Coarsening.
- Detecting Extreme Counterfactuals.
CEM Extensions I

- Many binary variables:

  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze

- Multicategory treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$

- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata

- Automating user choices

- Histogram bin size calculations, Estimated SATT error bound, Progressive Coarsening

- Detecting Extreme Counterfactuals
CEM Extensions I

- Many binary variables: $\Rightarrow$ coarsen sums of related vars
CEM Extensions I

- Many binary variables: \( \sim \) coarsen sums of related vars
- Missing Data and/or Measurement Error

1. Multiply Impute (missing) or Overimpute (measurement error)
2. Put missing observation in stratum where plurality of imputations fall
3. Pass on uncoarsened imputations to analysis stage
4. Use the usual MI combining rules to analyze

- Multicategory treatments: No modification necessary; keep all strata with \( \geq 1 \) unit having each value of \( T \)
- Blocking in Randomized Experiments: no modification needed; randomly assign \( T \) within CEM strata

- Automating user choices
  - Histogram bin size calculations,
  - Estimated SATT error bound,
  - Progressive Coarsening

- Detecting Extreme Counterfactuals
CEM Extensions I

- Many binary variables: \( \leadsto \) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
CEM Extensions I

- Many binary variables: \(\rightsquigarrow\) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
Many binary variables: \( \leadsto \) coarsen sums of related vars

Missing Data and/or Measurement Error
1. Multiply Impute (missing) or Overimpute (measurement error)
2. Put missing observation in stratum where plurality of imputations fall
3. Pass on uncoarsened imputations to analysis stage
CEM Extensions I

- Many binary variables: \(\sim\) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
CEM Extensions I

- Many binary variables: \(\sim\) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with \(\geq 1\) unit having each value of \(T\)
CEM Extensions I

- Many binary variables: \( \rightsquigarrow \) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with \( \geq 1 \) unit having each value of \( T \)
- Blocking in Randomized Experiments: no modification needed; randomly assign \( T \) within CEM strata
CEM Extensions I

- Many binary variables: \( \leadsto \) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with \( \geq 1 \) unit having each value of \( T \)
- Blocking in Randomized Experiments: no modification needed; randomly assign \( T \) within CEM strata
- Automating user choices
CEM Extensions I

- Many binary variables: \(\sim\) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with \(\geq 1\) unit having each value of \(T\)
- Blocking in Randomized Experiments: no modification needed; randomly assign \(T\) within CEM strata
- Automating user choices Histogram bin size calculations,
CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations, Estimated SATT error bound,
CEM Extensions I

- Many binary variables: \( \rightsquigarrow \) coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with \( \geq 1 \) unit having each value of \( T \)
- Blocking in Randomized Experiments: no modification needed; randomly assign \( T \) within CEM strata
- Automating user choices Histogram bin size calculations, \textit{Estimated} SATT error bound, Progressive Coarsening
CEM Extensions I

- Many binary variables: $\rightarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations, Estimated SATT error bound, Progressive Coarsening
- Detecting Extreme Counterfactuals
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties

- Inherent CEM properties if applied within CEM strata

- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn't be used in practice

- MDM: can apply within CEM strata

- Genetic Matching: can constrain results to CEM strata

- Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness

- Nonparametric Adjustments: can apply within CEM strata
Most commonly used methods:
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
Most commonly used methods:
- cannot be used to eliminate extrapolation region
- don’t possess most other CEM properties
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
Most commonly used methods:
- cannot be used to eliminate extrapolation region
- don’t possess most other CEM properties
- inherent CEM properties if applied within CEM strata

Propensity Score matching:
- CEM strata can bound bias in PSM
- Probably shouldn’t be used in practice

MDM: can apply within CEM strata
Genetic Matching: can constrain results to CEM strata
Synthetic Matching, or Robins’ weights: CEM can identify region to apply weights, increasing efficiency/robustness

Nonparametric Adjustments: can apply within CEM strata
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata

- Propensity Score matching:
  - CEM strata can bound bias in PSM
CEM Extensions II: Improving Existing Matching Methods

- **Most commonly used methods:**
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata

- **Propensity Score matching:**
  - CEM strata can bound bias in PSM
  - Probably shouldn’t be used in practice
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata

- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn’t be used in practice

- MDM: can apply within CEM strata
CEM Extensions II: Improving Existing Matching Methods

- **Most commonly used methods:**
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata

- **Propensity Score matching:**
  - CEM strata can bound bias in PSM
  - Probably shouldn’t be used in practice

- **MDM:** can apply within CEM strata

- **Genetic Matching:** can constrain results to CEM strata
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn’t be used in practice
- MDM: can apply within CEM strata
- Genetic Matching: can constrain results to CEM strata
- Synthetic Matching, or Robins’ weights: CEM can identify region to apply weights, increasing efficiency/robustness
CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don’t possess most other CEM properties
  - inherent CEM properties if applied within CEM strata

- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn’t be used in practice

- MDM: can apply within CEM strata

- Genetic Matching: can constrain results to CEM strata

- Synthetic Matching, or Robins’ weights: CEM can identify region to apply weights, increasing efficiency/robustness

- Nonparametric Adjustments: can apply within CEM strata
Part 3 (of 3)

The Matching Frontier
The Matching Frontier
The Matching Frontier

![Graph showing the matching frontier with two points labeled as Result #1 and Result #2. The x-axis represents the number of units pruned, ranging from low variance to high variance. The y-axis represents the imbalance, ranging from less biased to more biased.]}
The Matching Frontier

Number of Units Pruned

Imbalance

Low variance  High variance

Less biased  More biased

Result #1

Result #2

Result #3

Number of Units Pruned
The Matching Frontier

Number of Units Pruned

Imbalance

Low variance High variance

Less biased More biased

Result #1

Result #2

Result #3

Result #4
The Matching Frontier

The diagram illustrates the relationship between the number of units pruned, imbalance, and variance. The points labeled as Result #1, #2, #3, and #4 are plotted on the graph, showing how different levels of variance and bias are affected by the number of units pruned.
The Matching Frontier

Theoretical frontier (optimal)

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

Number of Units Pruned

Imbalance

Low variance  High variance

Less biased  More biased
The Matching Frontier

- Number of Units Pruned
- Imbalance
- Low variance
- High variance
- Less biased
- More biased

Results:
- 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 $(N \binom{n}{N})$ subsets of rows of $X_0$

Choose subset with lowest imbalance

Evaluations needed to compute the entire frontier:

$(N \binom{n}{N})$ evaluations for each sample size $n = N, N-1, ..., 1$

The combination is the (gargantuan) "power set"

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

$\Rightarrow$ It's hard to calculate!

We develop algorithms for the (optimal) frontier which:

- runs very fast
- operate as "greedy" but we prove are optimal
- do not require evaluating every subset
- work with very large data sets

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

- Consider 1 point on the SATT frontier:
How hard is the frontier to calculate?

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

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

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

$\Rightarrow$ It's hard to calculate!

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

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

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

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

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

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

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

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

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

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

- It’s hard to calculate!

- We develop algorithms for the (optimal) frontier which:
  - Runs very fast
  - Operate as “greedy” but we prove are optimal
  - Do not require evaluating every subset
  - Work with very large data sets
  - Is the exact frontier (no approximation or estimation)

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

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

- Evaluations needed to compute the entire frontier:

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

  The combination is the (gargantuan) “power set”

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

  $\Rightarrow$ It’s hard to calculate!

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

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

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

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

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

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

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

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

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

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

- Runs very fast
- Operate as “greedy” but we prove are optimal
- Do not require evaluating every subset
- Work with very large data sets
- Is the exact frontier (no approximation or estimation)

\[ \Rightarrow \text{It's easy to calculate!} \]
How hard is the frontier to calculate?

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

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

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

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

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

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

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

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

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

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

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

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

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

- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - 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
  • \( \rightsquigarrow \) It’s **hard** to calculate!

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

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

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

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

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

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

To use, make 2 choices:

1. Quantity of interest: SATT (prune Cs only) or FSATT
2. Fixed- or variable-ratio matching

Result:

• Algorithm finds the whole frontier
• Simple to use (free easy software available)
• No need to choose or use a matching method
• All solutions are optimal
• No iteration or diagnostics required
• No cherry picking possible; you see everything optimal
The Matching Frontier Algorithm

- To use, make 2 choices:
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

- Result:
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

- Result:
  - Algorithm finds the whole frontier
The Matching Frontier Algorithm

To use, make 2 choices:

1. Quantity of interest: SATT (prune Cs only) or FSATT
2. Fixed- or variable-ratio matching

Result:

- Algorithm finds the whole frontier
- Simple to use (free easy software available)
The Matching Frontier Algorithm

• To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

• Result:
  • Algorithm finds the whole frontier
  • Simple to use (free easy software available)
  • No need to choose or use a matching method
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method
  - All solutions are optimal
The Matching Frontier Algorithm

• To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

• Result:
  • Algorithm finds the whole frontier
  • Simple to use (free easy software available)
  • No need to choose or use a matching method
  • All solutions are optimal
  • No iteration or diagnostics required
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible; you see everything optimal
• 185 Ts; pruning most 16,252 Cs won’t increase variance much
• Huge bias-variance trade-off after pruning most Cs
• Estimates converge to experiment after removing bias
• No mysteries: basis of inference clearly revealed
Constructing the FSATT Mahalanobis Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0 -0.5 0.0 0.5 1.0

Treated
Control
Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0 0.1 0.2 0.3 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

- 0
- 5
- 10
- 15
- 20

- 0.0
- 0.1
- 0.2
- 0.3
- 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

-1.0 -0.5 0.0 0.5 1.0

Covariate 1

-1.0 -0.5 0.0 0.5 1.0

Covariate 2

0 5 10 15 20

Number of Observations Dropped

Average Mahalanobis Discrepancy

0.0 0.1 0.2 0.3 0.4

0.0 0.1 0.2 0.3 0.4

Treated
Control
Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy

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

0 5 10 15 20

0.0 0.1 0.2 0.3 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariates 1 and 2 are plotted on the graph.
- Data points represent treated, control, and next to remove observations.

Frontier

- Graph shows the average Mahalanobis discrepancy against the number of observations dropped.
- The frontier curve illustrates how the Mahalanobis discrepancy decreases as more observations are dropped.
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

○ Treated
○ Control
○ Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Treated
- Control
- Next to remove

Frontier

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

Remaining Data

Costariate 1
Costariate 2
-1.0 -0.5 0.0 0.5 1.0
-1.0
-0.5
0.0
0.5
1.0
●
●
●
●
●
●

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy

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

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1 vs. Covariate 2
- Red dots: Treated
- Blue dots: Control
- Blue circle: Next to remove

Frontier

- Average Mahalanobis Discrepancy vs. Number of Observations Dropped
- Red dots: Frontier points
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Treated
Control
Next to remove

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
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

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20
0.0 0.1 0.2 0.3 0.4

-0.5 0.0 0.5 1.0
-1.0
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

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

Remaining Data

Frontier

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

Short version:

<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>
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins

2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins

2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins

2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

**Short version:**

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units
Conclusions about Matching

• An excellent method of improving causal inferences
• Helps reduce imbalance, model dependence, researcher discretion, bias

• Propensity Score Matching: low standards, dangerous in practice; DNR
• Most other methods of matching: excellent
  • Coarsened Exact Matching: especially powerful; simple to understand and use
  • Matching Frontier: Automates what’s left to automate

• In applications, focus on the substance:
  • All observational methods must assume ignorability
  • Automated methods to choose variables: Insufficient
  • Claims of invariance: Always fail
  • Trying to be invariant to the substance: be wary of methods claiming to be invariant to what you know!
Conclusions about Matching

  
- Propensity Score Matching: low standards, dangerous in practice; DNR.
  
- Most other methods of matching: excellent.
  
- Coarsened Exact Matching: especially powerful; simple to understand and use.
  
- Matching Frontier: Automates what's left to automate.

In applications, focus on the substance:

- All observational methods must assume ignorability.
  
- Automated methods to choose variables: Insufficient.
  
- Claims of invariance: Always fail.
  
- Trying to be invariant to the substance: be wary of methods claiming to be invariant to what you know!
Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences

- Propensity Score Matching: low standards, dangerous in practice; DNR

- Most other methods of matching: excellent
  - Coarsened Exact Matching: especially powerful; simple to understand and use
  - Matching Frontier: Automates what's left to automate

In applications, focus on the substance:

- All observational methods must assume ignorability
- Automated methods to choose variables: Insufficient
- Claims of invariance: Always fail
- Trying to be invariant to the substance: be wary of methods claiming to be invariant to what you know!
Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
Conclusions about Matching

• Matching:
  • An excellent method of improving causal inferences
  • Helps reduce: imbalance, model dependence, researcher discretion, bias

• Propensity Score Matching: low standards, dangerous in practice; DNR

• Most other methods of matching: excellent

• Coarsened Exact Matching: especially powerful; simple to understand and use
Conclusions about Matching

- **Matching:**
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias

- **Propensity Score Matching:** low standards, dangerous in practice; DNR

- **Most other methods of matching:** excellent

- **Coarsened Exact Matching:** especially powerful; simple to understand and use

- **Matching Frontier:** Automates what’s left to automate
Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias

- Propensity Score Matching: low standards, dangerous in practice; DNR

- Most other methods of matching: excellent

- Coarsened Exact Matching: especially powerful; simple to understand and use

- Matching Frontier: Automates what’s left to automate

- In applications, focus on the substance:
Conclusions about Matching

- **Matching:**
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- **Propensity Score Matching:** low standards, dangerous in practice; DNR
- **Most other methods of matching:** excellent
- **Coarsened Exact Matching:** especially powerful; simple to understand and use
- **Matching Frontier:** Automates what’s left to automate
- **In applications, focus on the substance:**
  - All observational methods must assume ignorability
Conclusions about Matching

• Matching:
  • An excellent method of improving causal inferences
  • Helps reduce: imbalance, model dependence, researcher discretion, bias

• Propensity Score Matching: low standards, dangerous in practice; DNR

• Most other methods of matching: excellent

• Coarsened Exact Matching: especially powerful; simple to understand and use

• Matching Frontier: Automates what’s left to automate

• In applications, focus on the substance:
  • All observational methods must assume ignorability
  • Automated methods to choose variables: Insufficient
Conclusions about Matching

- **Matching:**
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias

- **Propensity Score Matching:** low standards, dangerous in practice; DNR

- **Most other methods of matching:** excellent

- **Coarsened Exact Matching:** especially powerful; simple to understand and use

- **Matching Frontier:** Automates what’s left to automate

- **In applications, focus on the substance:**
  - All observational methods must assume ignorability
  - Automated methods to choose variables: Insufficient
  - Claims of invariance: Always fail
Conclusions about Matching

• Matching:
  • An excellent method of improving causal inferences
  • Helps reduce: imbalance, model dependence, researcher discretion, bias

• Propensity Score Matching: low standards, dangerous in practice; DNR

• Most other methods of matching: excellent

• Coarsened Exact Matching: especially powerful; simple to understand and use

• Matching Frontier: Automates what’s left to automate

• In applications, focus on the substance:
  • All observational methods must assume ignorability
  • Automated methods to choose variables: Insufficient
  • Claims of invariance: Always fail
  • Trying to be invariant to the substance: be wary of methods claiming to be invariant to what you know!
Part 4 (of 3), :-)

Matching Theories of Inference (in one slide)
Assumptions to Justify Current Practice

Existing Theory of Inference

- **Framework**: simple random sampling from a population
- **Exact matching**: Rarely possible; but would make estimation easy
- **Assumptions**:
  - **Unconfoundedness**: $T \perp Y(0)|X$ (Healthy & unhealthy get meds)
  - **Common support**: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)

Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference

- **Framework**: stratified random sampling from a population
- **Define $A$**: a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- **We already know and use these procedures**: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- **Assumptions**:
  - **Set-wide Unconfoundedness**: $T \perp Y(0)|A$
  - **Set-wide Common support**: $\Pr(T = 1|A) < 1$
- **Fits all common matching methods & practices; no asymptotics**
- **Easy extensions for**: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) | X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 | X) < 1$, $T = 0, 1$ are both possible
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) | A$
  - Set-wide Common support: $\Pr(T = 1 | A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

**Existing Theory of Inference:** Stop What You’re Doing!

- **Framework:** simple random sampling from a population
- **Exact matching:** Rarely possible; but would make estimation easy
- **Assumptions:**
  - **Unconfoundedness:** $T \perp Y(0) | X$ (Healthy & unhealthy get meds)
  - **Common support:** $\Pr(T = 1 | X) < 1$ (T = 0, 1 are both possible)
- **Approximate matching (bias correction, new variance estimation):** common, but all current practices would have to change

**Alternative Theory of Inference:** It’s Gonna be OK!

- **Framework:** stratified random sampling from a population
- **Define $A$:** a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- **We already know and use these procedures:** Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- **Assumptions:**
  - **Set-wide Unconfoundedness:** $T \perp Y(0) | A$
  - **Set-wide Common support:** $\Pr(T = 1 | A) < 1$
- **Fits all common matching methods & practices; no asymptotics**
- **Easy extensions for:** multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

**Existing Theory of Inference:** Stop What You’re Doing!

- Framework: *simple random sampling* from a population

**Alternative Theory of Inference:** It’s Gonna be OK!

- Framework: *stratified random sampling* from a population
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) | A$
  - Set-wide Common support: $\Pr(T = 1 | A) < 1$ (T = 0, 1 are both possible)
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

**Existing Theory of Inference:** Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:

**Alternative Theory of Inference:** It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) | A$
  - Set-wide Common support: $\Pr(T = 1 | A) < 1$ (both possible)
  - Fits all common matching methods & practices; no asymptotics
  - Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $\Pr(T = 1 \mid A) < 1$
  - Fits all common matching methods & practices; no asymptotics
  - Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $\Pr(T = 1 \mid A) < 1$
  - Fits all common matching methods & practices; no asymptotics
  - Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)

- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $Pr(T = 1 \mid A) < 1$ ($T = 0, 1$ are both possible)
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: **simple random sampling** from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness:* \( T \perp Y(0) \mid X \) (Healthy & unhealthy get meds)
  - *Common support:* \( \Pr(T = 1 \mid X) < 1 \) (\( T = 0, 1 \) are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: **stratified random sampling** from a population
- Define \( A \): a stratum in a partition of the product space of \( X \) (“continuous” variables have natural breakpoints)
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: **simple random sampling** from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: **stratified random sampling** from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $\Pr(T = 1 \mid A) < 1$
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don’t match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $\Pr(T = 1 \mid A) < 1$
- Fits all common matching methods & practices; no asymptotics
Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You’re Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - Common support: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It’s Gonna be OK!

- Framework: stratified random sampling from a population
- Define $A$: a stratum in a partition of the product space of $X$ (“continuous” variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - Set-wide Unconfoundedness: $T \perp Y(0) \mid A$
  - Set-wide Common support: $\Pr(T = 1 \mid A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small