The Balance-Sample Size Frontier in Matching Methods for Causal Inference\textsuperscript{1}

Gary King\textsuperscript{2}

Institute for Quantitative Social Science
Harvard University

(Talk at the University of Michigan, 1/24/2014)

\textsuperscript{1}Joint work with Christopher Lucas and Richard Nielsen
\textsuperscript{2}GaryKing.org.
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)

- Problem: Matching prunes \( n \) to improve imbalance, but
  - Some: set \( n \) and don't guarantee imbalance
  - Others: set imbalance and don't guarantee \( n \)
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (\( n \)-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier

- Side point:
  - Problem: Propensity score matching increases imbalance!
  - Solution: Not an issue with other methods or our approach
Overview

- Problem: Model dependence (review)

- Solution: Matching to reduce model dependence (review)

- Problem: Matching prunes $n$ to improve imbalance, but:
  - Some: set $n$ and don't guarantee imbalance
  - Others: set imbalance and don't guarantee $n$
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks

- Solution: easier & more powerful
  - Estimate the ($n$-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier

Side point:

- Problem: Propensity score matching increases imbalance!
- Solution: Not an issue with other methods or our approach
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching prunes $n$ to improve imbalance, but...
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching prunes $n$ to improve imbalance, but
  • Some: set $n$ and don’t guarantee imbalance
• Solution: easier & more powerful
  • Estimate the ($n$-imbalance) “matching frontier”
  • Imbalance metric choice defines the frontier

Side point:
• Problem: Propensity score matching increases imbalance!
• Solution: Not an issue with other methods or our approach
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching prunes $n$ to improve imbalance, but
  • Some: set $n$ and don’t guarantee imbalance
  • Others: set imbalance and don’t guarantee $n$
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes $n$ to improve imbalance, but
  - Some: set $n$ and don’t guarantee imbalance
  - Others: set imbalance and don’t guarantee $n$
  - Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks

Estimate the ($n$-imbalance) “matching frontier”

Imbalance metric choice defines the frontier

Side point:
- Problem: Propensity score matching increases imbalance!
- Solution: Not an issue with other methods or our approach
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching prunes $n$ to improve imbalance, but
  • Some: set $n$ and don’t guarantee imbalance
  • Others: set imbalance and don’t guarantee $n$
  • Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
• Solution: easier & more powerful
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes \( n \) to improve imbalance, but
  - Some: set \( n \) and don’t guarantee imbalance
  - Others: set imbalance and don’t guarantee \( n \)
  - Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the \((n\text{-imbalance})\) “matching frontier”
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes $n$ to improve imbalance, but
  - Some: set $n$ and don’t guarantee imbalance
  - Others: set imbalance and don’t guarantee $n$
  - Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the ($n$-imbalance) “matching frontier”
  - Imbalance metric choice defines the frontier
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes $n$ to improve imbalance, but
  - Some: set $n$ and don’t guarantee imbalance
  - Others: set imbalance and don’t guarantee $n$
  - Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the ($n$-imbalance) “matching frontier”
  - Imbalance metric choice defines the frontier
- Side point:
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes $n$ to improve imbalance, but
  - Some: set $n$ and don’t guarantee imbalance
  - Others: set imbalance and don’t guarantee $n$
  - Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the ($n$-imbalance) “matching frontier”
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: Propensity score matching increases imbalance!
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching prunes $n$ to improve imbalance, but
  • Some: set $n$ and don’t guarantee imbalance
  • Others: set imbalance and don’t guarantee $n$
  • Plus: Matching methods optimize a different “imbalance” than recommended post-hoc checks
• Solution: easier & more powerful
  • Estimate the ($n$-imbalance) “matching frontier”
  • Imbalance metric choice defines the frontier
• Side point:
  • Problem: Propensity score matching increases imbalance!
  • Solution: Not an issue with other methods or our approach
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>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 R²</td>
<td>.423</td>
<td></td>
</tr>
</tbody>
</table>
Doyle and Sambanis: Model Dependence

In Sample Fit

Counterfactual Prediction
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching within the Interpolation Region

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

Matching reduces model dependence, bias, and variance
How Matching Works

- Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- Estimation:
  - Treatment Effect for treated ($T_i = 1$): $TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)$ = observed - unobserved

- Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)

- Prune unmatched units to improve balance (so $X$ is unimportant)

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated: $SATT = \frac{\sum_{i \in \{T_i = 1\}} (TE_i)}{\sum_{i \in \{T_i = 1\}}}$
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

- Notation:
  - \( Y_i \): Dependent variable
  - \( T_i \): Treatment variable (0/1, or more general)
  - \( X_i \): Pre-treatment covariates

- Estimation
  - Treatment Effect for treated \( (T_i = 1) \) observation:
    \[
    TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}
    \]

- Estimate \( Y_i(T_i = 0) \) with \( Y_j \) from matched \( (X_i \approx X_j) \) controls
  \[
  \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
  \]

- Prune unmatched units to improve balance (so \( X \) is unimportant)

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \frac{1}{\sum_i \mathbb{1}(T_i = 1)} \frac{1}{\sum_i \mathbb{1}(T_i = 1)} \sum_{i \in \{T_i = 1\}} (TE_i)
     \]
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

• Notation:
  \( Y_i \) Dependent variable
How Matching Works

- Notation:
  \( Y_i \) Dependent variable
  \( T_i \) Treatment variable (0/1, or more general)

- Estimation
  Treatment Effect for treated (\( T_i = 1 \)) observation:
  \[
  TE_i = Y_i (T_i = 1) - Y_i (T_i = 0) = \text{observed} - \text{unobserved}
  \]

- Estimate \( Y_i (T_i = 0) \) with \( Y_j \) from matched (\( X_i \approx X_j \)) controls
  \[ \hat{Y}_i (T_i = 0) = Y_j (T_i = 0) \] (or a model)

- Prune unmatched units to improve balance (so \( X \) is unimportant)

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
  \[
  SATT = \text{mean}_{i \in \{T_i = 1\}} (TE_i)
  \]
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

- Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation:
    \[ TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) \]
    - Observed - Unobserved
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
    \[ \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)} \]
  - Prune unmatched units to improve balance (so $X$ is unimportant)

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
    \[ SATT = \text{mean}_{i \in \{ T_i = 1 \}}(TE_i) \]
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

• Notation:
  \( Y_i \) Dependent variable
  \( T_i \) Treatment variable (0/1, or more general)
  \( X_i \) Pre-treatment covariates

• Estimation

  Treatment Effect for treated (\( T_i = 1 \)) observation
  \[ TE_i = Y_i(\ T_i = 1) - Y_i(\ T_i = 0) \]
  - observed - unobserved

  Estimate \( Y_i(\ T_i = 0) \) with \( Y_j \) from matched (\( X_i \approx X_j \)) controls
  \[ \hat{Y}_i(\ T_i = 0) = Y_j(\ T_i = 0) \] (or a model)

• Prune unmatched units to improve balance (so \( X \) is unimportant)

• Quantities of Interest:
  1. \( SATT \): Sample Average Treatment effect on the Treated:
     \[ SATT = \text{mean}_{i \in \{ T_i = 1 \}} (TE_i) \]
  2. \( FSATT \): Feasible Average Treatment effect on the Treated
How Matching Works

• Notation:
  \( Y_i \) Dependent variable
  \( T_i \) Treatment variable (0/1, or more general)
  \( X_i \) Pre-treatment covariates

• Estimation
  • **Treatment Effect** for treated \((T_i = 1)\) observation \(i\):

\[
\text{TE}_i = Y_i(\text{T}_i = 1) - Y_i(\text{T}_i = 0)
\]

• Estimate \( Y_i(\text{T}_i = 0) \) with \( Y_j \) from matched \((X_i \approx X_j)\) controls

• Prune unmatched units to improve balance (so \(X\) is unimportant)

• Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(\text{TE}_i)
     \]
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

• Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

• Estimation
  - Treatment Effect for treated ($T_i = 1$) observation $i$:
    \[
    TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)
    \]
How Matching Works

- **Notation:**
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- **Estimation**
  - **Treatment Effect** for treated ($T_i = 1$) observation $i$:
    \[
    TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}
    \]
How Matching Works

- **Notation:**
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- **Estimation**
  - **Treatment Effect** for treated ($T_i = 1$) observation $i$:
    \[ TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) \]
    \[ = \text{observed} - \text{unobserved} \]
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
    \[ \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \] (or a model)

- Prune unmatched units to improve balance (so $X$ is unimportant)

- **Quantities of Interest:**
  1. SATT: Sample Average Treatment effect on the Treated:
     \[ SATT = \text{mean}_{i \in \{T_i = 1\}}(TE_i) \]
  2. FSATT: Feasible Average Treatment effect on the Treated
How Matching Works

- **Notation:**
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- **Estimation**
  - **Treatment Effect** for treated ($T_i = 1$) observation $i$:
    \[
    TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)
    = \text{observed} - \text{unobserved}
    \]
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
    \]
  - Prune unmatched units to improve balance (so $X$ is unimportant)

- **Quantities of Interest:**
  1. **SATT:** Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \frac{1}{i \in \{T_i = 1\}} \sum TE_i
     \]
  2. **FSATT:** Feasible Average Treatment effect on the Treated
     \[
     \text{FSATT} = \text{observed} - \text{unobserved}
     \]
How Matching Works

- Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation $i$:
    \[
    TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)
    \]
    = observed − unobserved
  
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
    \]
  - Prune unmatched units to improve balance (so $X$ is unimportant)
  
- Quantities of Interest:
How Matching Works

• Notation:
  \( Y_i \) Dependent variable
  \( T_i \) Treatment variable (0/1, or more general)
  \( X_i \) Pre-treatment covariates

• Estimation
  • Treatment Effect for treated \((T_i = 1)\) observation \(i\):
    \[
    \text{TE}_i = Y_i( T_i = 1) - Y_i( T_i = 0)
    \]
    = observed \(-\) unobserved
  
  • Estimate \( Y_i( T_i = 0) \) with \( Y_j \) from matched \((X_i \approx X_j)\) controls
    \( \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \) (or a model)
  
  • Prune unmatched units to improve balance (so \(X\) is unimportant)

• Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
    \[
    \text{SATT} = \text{mean}_{i \in \{T_i = 1\}} \left( \text{TE}_i \right)
    \]
How Matching Works

• Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1, or more general)
  - $X_i$: Pre-treatment covariates

• Estimation
  - Treatment Effect for treated ($T_i = 1$) observation $i$:
    \[
    \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)
    \]
    \[
    = \text{observed} - \text{unobserved}
    \]
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
    \]
  - Prune unmatched units to improve balance (so $X$ is unimportant)

• Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
    \[
    \text{SATT} = \text{mean}_{i \in \{T_i = 1\}} (\text{TE}_i)
    \]
  2. FSATT: Feasible Average Treatment effect on the Treated
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching

1. **Preprocess** (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$

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

3. **Checking** Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching

1. **Preprocess** (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)^\prime S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Method 1: Mahalanobis Distance Matching

1. **Preprocess** (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Method 1: Mahalanobis Distance Matching

1. **Preprocess (Matching)**
   - Distance\( (X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > *caliper*

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

3. **Checking** Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching

1. **Preprocess (Matching)**
   - Distance \( (X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > *caliper*

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

3. **Checking** Measure imbalance, tweak, repeat, . . .
Mahalanobis Distance Matching

Age

Education (years)
Mahalanobis Distance Matching

Education (years)

Age

12 14 16 18 20 22 24 26 28
Mahalanobis Distance Matching
Mahalanobis Distance Matching

Education (years)

Age

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

T T

T T

TT TT T TT

TTT TT

T

T TT

C

C

C C

CC

C

C

C CCC

C

C

CCCC C CCC C

CCCC

C CCCC C
Mahalanobis Distance Matching

Age
12 14 16 18 20 22 24 26 28
20
30
40
50
60
70
80

Education (years)
Mahalanobis Distance Matching

Age
Education (years)

12 14 16 18 20 22 24 26 28

20
30
40
50
60
70
80

T T
T T
TT TT T TT
TTT TT
T
T TT
C
C
C C
CC
C
C
C CCC
C
C
CC C C C
C
Method 2: Propensity Score Matching

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

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, . . .
Method 2: Propensity Score Matching

1. Preprocess (Matching)

2. Estimation  Difference in means or a model
3. Checking  Measure imbalance, tweak, repeat, . . .
Method 2: Propensity Score Matching

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

2. Estimation
   Difference in means or a model

3. Checking
   Measure imbalance, tweak, repeat, . . .
Method 2: Propensity Score Matching

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

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...
Method 2: Propensity Score Matching

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

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, ...
Method 2: Propensity Score Matching

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

   - Distance($X_i$, $X_j$) = $|\pi_i - \pi_j|$

   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation  Difference in means or a model
3. Checking  Measure imbalance, tweak, repeat, ...
Method 2: Propensity Score Matching

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

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, ...
Method 2: Propensity Score Matching

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

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, ...
Propensity Score Matching

![Propensity Score Matching Graph](image_url)
Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - 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 for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
   - Sort observations into strata, each with unique values of $C(X)$
   - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. Estimation
   - Difference in means or a model
     - Need to weight controls in each stratum to equal treateds
     - Can apply other matching methods within CEM strata (inherit CEM’s properties)

3. Checking
   - Determine matched sample size, tweak, repeat, . . .
     - Easier, but still iterative
Method 3: Coarsened Exact Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

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

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

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

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

1. **Preprocess (Matching)**
   - 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 for a histogram

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   • 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 for a histogram
   • Apply exact matching to the coarsened $X$, $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...
Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - 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 for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

1. **Preprocess (Matching)**
   - **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 for a histogram
   - **Apply exact matching** to the coarsened $X, C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 3: Coarsened Exact Matching

1. **Preprocess (Matching)**
   - **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 for a histogram
   - **Apply exact matching** to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - **Pass on original (uncoarsened) units** except those pruned

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

1. **Preprocess (Matching)**
   - 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 for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 3: Coarsened Exact Matching

1. **Preprocess** (Matching)
   - 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 for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. **Estimation** Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM’s properties)

3. **Checking** Determine matched sample size, tweak, repeat, …
Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - 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 for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned

2. Estimation
   - Difference in means or a model
     - Need to weight controls in each stratum to equal treateds
     - Can apply other matching methods within CEM strata (inherit CEM’s properties)

3. Checking
   - Determine matched sample size, tweak, repeat, …
   - Easier, but still iterative
Coarsened Exact Matching
Coarsened Exact Matching

Age

12 14 16 18 20 22 24 26 28

Education

20
30
40
50
60
70
80

CCC C
CC CC
C CC C CCC CCCC CCC CC CCC CCCCCC
C CCC CC C
T T
T T
TT TT T TT
TTT TT
T
T TT

27 / 55
Coarsened Exact Matching

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

Education
HS BA MA PhD 2nd PhD

Senior Discounts

Don't trust anyone over 30
The Big 40
Coarsened Exact Matching

Age

Education

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

Coarsened Exact Matching
The Matching Frontier

Bias-Variance trade off ⇝ Imbalance

Frontier = matched dataset with lowest imbalance for each

To use, make 3 choices:

1. Imbalance metric, e.g.:
   - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
   - Difference of multivariate histograms ($L_1$):

2. Quantity of interest: SATT (prune Cs only) or FSATT

3. Fixed- or variable-ratio matching

Result:

- Simple to use
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible
The Matching Frontier

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

- Bias-Variance trade off $\rightsimeq$ Imbalance-$n$ Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching

- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible
The Matching Frontier

• Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$

• To use, make 3 choices:
  1. Imbalance metric, e.g.:
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  - Frontier $=$ matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible
The Matching Frontier

- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
The Matching Frontier

- Bias-Variance trade off \(\sim\) Imbalance-\(n\) Trade Off
  Frontier = matched dataset with lowest imbalance for each \(n\)
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms \((L_1)\):
  2. Quantity of interest: SATT (prune Cs only) or FSATT

---

33 / 55
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
The Matching Frontier

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

- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching

- Result:
The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each
       unit to the closest in the other treatment regime)
     - Difference of multivariate histograms ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
The Matching Frontier

- Bias-Variance trade off \( \sim \) Imbalance-\( n \) Trade Off
  Frontier = matched dataset with lowest imbalance for each \( n \)
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (\( L_1 \))
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required
The Matching Frontier

- Bias-Variance trade off \( \sim \) Imbalance-\( n \) Trade Off
  
  Frontier = matched dataset with lowest imbalance for each \( n \)

- To use, make 3 choices:
  
  1. Imbalance metric, e.g.:
     
     - Average Mahalanobis Distance (average distance from each
       unit to the closest in the other treatment regime)
     
     - Difference of multivariate histograms \( (L_1) \):
  
  2. Quantity of interest: SATT (prune Cs only) or FSATT

  3. Fixed- or variable-ratio matching

- Result:
  
  - Simple to use
  
  - All solutions are optimal
  
  - No iteration or diagnostics required
  
  - No cherry picking possible
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 \((N \times k)\) control matrix \(X_0\)

\[\binom{N}{n}\] 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!

\(\Rightarrow\) It's hard to calculate!

We develop new algorithms for several frontiers which:

- run very fast
- 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:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)

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

The combination is the (gargantuan) “power set”

- e.g., \(N > 300\) requires more imbalance evaluations than elementary particles in the universe!

\(\Rightarrow\) It’s hard to calculate!

- We develop new algorithms for several frontiers which:
  - run very fast
  - 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:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose the (or a) subset with the lowest imbalance

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

- \( \rightarrow \) It's hard to calculate!

- We develop new algorithms for several frontiers which:
  - run very fast
  - 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:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose the (or a) subset with the lowest imbalance

- Evaluations needed to compute the entire frontier:
  - \((N \times k)\) 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 new algorithms for several frontiers which:
  - run very fast
  - 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:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose the (or a) subset with the lowest imbalance

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

\(\Rightarrow\) It's hard to calculate!

\(\Rightarrow\) We develop new algorithms for several frontiers which:
  - Run very fast
  - 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:
  • Start with \((N \times k)\) control matrix \(X_0\)
  • Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  • Choose the (or a) subset with the 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”

\(\Rightarrow\) It's hard to calculate!

We develop new algorithms for several frontiers which:
  • run very fast
  • 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:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose the (or a) subset with the 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!

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

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

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

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

• We develop new algorithms for several frontiers which:
How hard is the frontier to calculate?

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

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

• We develop new algorithms for several frontiers which:
  • run very fast
How hard is the frontier to calculate?

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

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

- We develop new algorithms for several frontiers which:
  - run very fast
  - do not require evaluating every subset
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with \((N \times k)\) control matrix \(X_0\)
  - Calculate imbalance for all \(\binom{N}{n}\) subsets of rows of \(X_0\)
  - Choose the (or a) subset with the 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 new algorithms for several frontiers which:
  - run very fast
  - 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 \( (N \times k) \) control matrix \( X_0 \)
  - Calculate imbalance for all \( \binom{N}{n} \) subsets of rows of \( X_0 \)
  - Choose the (or a) subset with the 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 new algorithms for several frontiers which:
  - run very fast
  - do not require evaluating every subset
  - work with very large data sets
  - \( \rightsquigarrow \) It’s **easy** to calculate!
Job Training Data: Frontier and Causal Estimates
Job Training Data: Frontier and Causal Estimates
185 Ts; pruning most 16,252 Cs won’t increase variance much
Job Training Data: Frontier and Causal Estimates

- 185 Ts; pruning most 16,252 Cs won’t increase variance much
- Huge bias-variance trade-off after most are pruned
Job Training Data: Frontier and Causal Estimates

- 185 Ts; pruning most 16,252 Cs won’t increase variance much
- Huge bias-variance trade-off after most are pruned
- Estimates converge to experiment after removing bias
Job Training Data: Frontier and Causal Estimates

- 185 Ts; pruning most 16,252 Cs won’t increase variance much
- Huge bias-variance trade-off after most are pruned
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed
Aid Shocks Data: Frontier and Causal Estimates
Aid Shocks Data: Frontier and Causal Estimates
Aid Shocks Data: Frontier and Causal Estimates

- Frontier is nearly linear (left)
Aid Shocks Data: Frontier and Causal Estimates

- Frontier is nearly linear (left)
- Causal effects have big jumps (right)
Aid Shocks Data: Frontier and Causal Estimates

- Frontier is nearly linear (left)
- Causal effects have big jumps (right)
- More difficult inferential task
Aids Shocks: Change in Quantity of Interest
Aids Shocks: Change in Quantity of Interest

Number of Observations Pruned

Standardized Mean Value

0 500 1000 1500

Ethnic Frac.
Cold War
Religious Frac.
Infant Mortality
ln GDP p.c.
Mountains
Human Rights
ln Population
Partial Autocracy
Partial Democracy
Bad Neighborhood
Instability
Factional Democracy
Full Democracy
Noncontiguous
Demonstrations
Riots
Oil
Strikes
Assassinations
Aids Shocks: Large Unit-Level Effects

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Case</th>
<th>Effect change</th>
<th>Remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia</td>
<td>1991</td>
<td>0</td>
<td>0.008 → 0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger</td>
<td>1994</td>
<td>1</td>
<td>0.015 → 0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho</td>
<td>1998</td>
<td>1</td>
<td>0.021 → 0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D’Ivoire</td>
<td>2002</td>
<td>1</td>
<td>0.011 → 0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea</td>
<td>2000</td>
<td>1</td>
<td>0.005 → 0.0</td>
<td>739</td>
</tr>
</tbody>
</table>

- High leverage points
- Cases with few substitutes
- Not model dependence (which matching helps with), but data dependence
## Aids Shocks: Large Unit-Level Effects

<table>
<thead>
<tr>
<th>Case</th>
<th>$T$</th>
<th>$Y$</th>
<th>Effect change</th>
<th>N remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia, 1991</td>
<td>1</td>
<td>0</td>
<td>0.008→0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger, 1994</td>
<td>0</td>
<td>1</td>
<td>0.015→0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho, 1998</td>
<td>1</td>
<td>1</td>
<td>0.021→0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D’Ivoire, 2002</td>
<td>1</td>
<td>1</td>
<td>0.011→0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea, 2000</td>
<td>1</td>
<td>1</td>
<td>0.005→0</td>
<td>739</td>
</tr>
</tbody>
</table>
## Aids Shocks: Large Unit-Level Effects

<table>
<thead>
<tr>
<th>Case</th>
<th>( T )</th>
<th>( Y )</th>
<th>Effect change</th>
<th>( N ) remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia, 1991</td>
<td>1</td>
<td>0</td>
<td>0.008→0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger, 1994</td>
<td>0</td>
<td>1</td>
<td>0.015→0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho, 1998</td>
<td>1</td>
<td>1</td>
<td>0.021→0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D’Ivoire, 2002</td>
<td>1</td>
<td>1</td>
<td>0.011→0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea, 2000</td>
<td>1</td>
<td>1</td>
<td>0.005→0</td>
<td>739</td>
</tr>
</tbody>
</table>

- **High leverage points**
### Aids Shocks: Large Unit-Level Effects

<table>
<thead>
<tr>
<th>Case</th>
<th>$T$</th>
<th>$Y$</th>
<th>Effect change</th>
<th>N remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia, 1991</td>
<td>1</td>
<td>0</td>
<td>0.008$\rightarrow$0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger, 1994</td>
<td>0</td>
<td>1</td>
<td>0.015$\rightarrow$0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho, 1998</td>
<td>1</td>
<td>1</td>
<td>0.021$\rightarrow$0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D’Ivoire, 2002</td>
<td>1</td>
<td>1</td>
<td>0.011$\rightarrow$0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea, 2000</td>
<td>1</td>
<td>1</td>
<td>0.005$\rightarrow$0</td>
<td>739</td>
</tr>
</tbody>
</table>

- High leverage points
- Cases with few substitutes
## Aids Shocks: Large Unit-Level Effects

<table>
<thead>
<tr>
<th>Case</th>
<th>T</th>
<th>Y</th>
<th>Effect change</th>
<th>N remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia, 1991</td>
<td>1</td>
<td>0</td>
<td>0.008→0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger, 1994</td>
<td>0</td>
<td>1</td>
<td>0.015→0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho, 1998</td>
<td>1</td>
<td>1</td>
<td>0.021→0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D’Ivoire, 2002</td>
<td>1</td>
<td>1</td>
<td>0.011→0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea, 2000</td>
<td>1</td>
<td>1</td>
<td>0.005→0</td>
<td>739</td>
</tr>
</tbody>
</table>

- High leverage points
- Cases with few substitutes
- Not model dependence (which matching helps with), but data dependence
Constructing the FSATT Mahalanobis Frontier
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2
-1.0 -0.5 0.0 0.5 1.0

Frontier

Number of Observations Dropped
Average Mahalanobis Discrepancy
0 5 10 15 20
0.0 0.1 0.2 0.3 0.4

○ Treated
○ Control
○ Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Covariate 1 values: -1.0, -0.5, 0.0, 0.5, 1.0

Covariate 2 values: -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

Treated
Control
Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

0 5 10 15 20

0.0

0.1

0.2

0.3

0.4

Treated

Control

Next to remove

Average Mahalanobis Discrepancy

Number of Observations Dropped

0 5 10 15 20

0.0

0.1

0.2

0.3

0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Covariate 1
- Covariate 2

○ Treated
○ Control
○ Next to remove

Average Mahalanobis Discrepancy

Number of Observations Dropped

-0.5 -0.0 0.5 1.0
-1.0 0.0 1.0 2.0
0.0 0.1 0.2 0.3 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1

Covariate 2

-1.0 −0.5 0.0 0.5 1.0

-1.0 −0.5 0.0 0.5 1.0

0 5 10 15 20

0.0

0.1

0.2

0.3

0.4

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Treated
- Control
- Next to remove

Frontier

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

Remaining Data

Frontier

Covariate 1

Covariate 2

Treated
● Control
○ Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

**Remaining Data**
- 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

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1
Covariate 2

-1.0 −0.5 0.0 0.5 1.0

-1.0
−0.5
0.0
0.5
1.0
●
●
●
●
●
●
●
●
●

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

○ Treated
□ Control
○ Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped

Covariate 1

Covariate 2
Constructing the FSATT Mahalanobis Frontier

Remaining Data

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

Frontier

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

Remaining Data

Frontier

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

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0

0.1

0.2

0.3

0.4

Number of Observations Dropped

0 5 10 15 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

○ Treated
○ Control
○ Next to remove

Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

Covariate 1

Covariate 2
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

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

- **Warning**: figure omits some details!
Constructing the FSATT Mahalanobis Frontier

Warning: figure omits some details!

- Very fast; works with any continuous imbalance metric
Constructing the L1/L2 SATT Frontier

![Bar Chart]

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

![Scatter Plot]

- Frequency of observations dropped:
  - Bin 1: 0.12
  - Bin 2: 0.10
  - Bin 3: 0.08
  - Bin 4: 0.06
  - Bin 5: 0.04
  - Bin 6: 0.02

- Number of observations dropped:
  - 0
  - 2
  - 4
  - 6
  - 8
  - 10
Constructing the L1/L2 SATT Frontier

![Graph showing frequency and L2 values for different bins. The graph compares Treatment and Control groups across bins 1 to 6. The L2 values range from 0 to 0.12, and the number of observations dropped range from 0 to 10.](image-url)
Constructing the L1/L2 SATT Frontier

Table:

<table>
<thead>
<tr>
<th>Bin</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>5</td>
</tr>
<tr>
<td>Bin2</td>
<td>6</td>
</tr>
<tr>
<td>Bin3</td>
<td>3</td>
</tr>
<tr>
<td>Bin4</td>
<td>3</td>
</tr>
<tr>
<td>Bin5</td>
<td>3</td>
</tr>
<tr>
<td>Bin6</td>
<td>1</td>
</tr>
</tbody>
</table>

Graph:

- x-axis: Number of Observations Dropped
- y-axis: L2
- Data points: (0, 0.12), (2, 0.10)

Legend:
- □ Treatment
- ■ Control

Note: The number of observations dropped in each bin is as follows: Bin1: 5, Bin2: 6, Bin3: 2, Bin4: 3, Bin5: 3, Bin6: 1.
Constructing the L1/L2 SATT Frontier

The diagram shows the frequency distribution of observations for different bins labeled as Bin1 to Bin6. The bars represent the number of observations dropped for Treatment and Control groups.

A scatter plot on the right side of the diagram illustrates the relationship between the number of observations dropped and a variable labeled L2.
Constructing the L1/L2 SATT Frontier

![Bar graph showing frequency distribution across bins for Treatment and Control groups.]

![Graph showing the number of observations dropped vs. L2 value.]

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

- **Number of Observations Dropped vs. L2 Value:**
  - L2 values range from 0.00 to 0.12
  - Number of observations dropped range from 0 to 10
  - The graph shows a downward trend indicating a decrease in L2 with an increase in the number of observations dropped.
Constructing the L1/L2 SATT Frontier

Frequency

- Bin 1: 4
- Bin 2: 6
- Bin 3: 3
- Bin 4: 3
- Bin 5: 2
- Bin 6: 1

Treatment vs. Control

Number of Observations Dropped

L2

0.00 0.02 0.04 0.06 0.08 0.10 0.12

0 2 4 6 8 10
Constructing the L1/L2 SATT Frontier

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

Graph shows the number of observations dropped vs. L2 with a decreasing trend.

Legend: [Color] Treatment, [Color] Control
Constructing the L1/L2 SATT Frontier

![Bar chart and line graph showing the frequency of observations in different bins for Treatment and Control groups, along with the number of observations dropped and their corresponding L2 values.](image-url)
Constructing the L1/L2 SATT Frontier

![Bar chart and line graph showing frequency and L2 values for different bins.]

- **Bar chart**:
  - Bins 1 to 6, with frequency counts.
  - Two bars for Treatment and Control.

- **Line graph**:
  - X-axis: Number of observations dropped.
  - Y-axis: L2 values.

- **Legend**:
  - Blue: Treatment.
  - Red: Control.

- **Example values**:
  - Bin 1: 4 (Treatment), 4 (Control).
  - Bin 2: 6 (Treatment), 6 (Control).
  - Bin 3: 2 (Treatment), 2 (Control).
  - Bin 4: 3 (Treatment), 3 (Control).
  - Bin 5: 2 (Treatment), 2 (Control).
  - Bin 6: 1 (Treatment), 2 (Control).

**Summary**:
- The chart illustrates the distribution of observations across different bins for Treatment and Control groups.
- The line graph shows a decreasing trend in L2 values as the number of observations dropped increases.

**Additional Note**:
- The notation L2 / 55 indicates a possible ratio or specific value related to the L2 metric.
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distribution for different bins.

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

![Graph showing L2 values against number of observations dropped.

- L2 values decrease as the number of observations dropped increases.

Legend:
- Blue bars: Treatment
- Red bars: Control

L2 / 55
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

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

- Warning: This figure omits some technical details too!
- Works very fast, even with very large data sets
Problems with PSM: Foreign Aid Shocks
King, Nielsen, Coberley, Pope, and Wells (2012)

Imbalance Metric

<table>
<thead>
<tr>
<th>Mahalanobis Discrepancy</th>
<th>L₁</th>
<th>Difference in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>published PSM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>published PSM with 1/4 sd caliper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Pruning</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Best Practices" PSM
MDM
CEM

Methods-specific frontiers (for methodological research only)
Problems with PSM: Healthways Data

King, Nielsen, Coberley, Pope, and Wells (2012)

Imbalance Metric

<table>
<thead>
<tr>
<th>Mahalanobis Discrepancy</th>
<th>L₁</th>
<th>Difference in Means</th>
</tr>
</thead>
</table>

- **Raw Data**
- **Random Pruning**
- "Best Practices" PSM
- PSM
- MDM
- CEM

Methods-specific frontiers (for methodological research only)
PSM Approximates Random Matching in Balanced Data

PSM Matches

CEM and MDM Matches
Conclusions
Conclusions

- The Matching Frontier
Conclusions

• The Matching Frontier
  • Fast; easy to use; no need to iterate
Conclusions

- **The Matching Frontier**
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
Conclusions

• The Matching Frontier
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric
Conclusions

• The Matching Frontier
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric
• Propensity score matching:
Conclusions

• **The Matching Frontier**
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric

• **Propensity score matching:**
  • The problem:
Conclusions

- **The Matching Frontier**
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

- **Propensity score matching:**
  - The problem:
    - Imbalance can be worse than original data
Conclusions

• **The Matching Frontier**
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

• **Propensity score matching:**
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
Conclusions

• **The Matching Frontier**
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric
• **Propensity score matching:**
  • The problem:
    • Imbalance can be worse than original data
    • Can increase imbalance when removing the worst matches
    • Approximates random matching in well-balanced data
      (Random matching increases imbalance)
Conclusions

• The Matching Frontier
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric

• Propensity score matching:
  • The problem:
    • Imbalance can be worse than original data
    • Can increase imbalance when removing the worst matches
    • Approximates random matching in well-balanced data
    (Random matching increases imbalance)
  • Implications:
Conclusions

- **The Matching Frontier**
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

- **Propensity score matching:**
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data
    (Random matching increases imbalance)
  - Implications:
    - Balance checking required
Conclusions

• The Matching Frontier
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric

• Propensity score matching:
  • The problem:
    • Imbalance can be worse than original data
    • Can increase imbalance when removing the worst matches
    • Approximates random matching in well-balanced data
      (Random matching increases imbalance)
  • Implications:
    • Balance checking required
    • Adjusting for potentially irrelevant covariates with PSM: mistake
Conclusions

- **The Matching Frontier**
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

- **Propensity score matching**
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data
      (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data
      (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates \textit{with PSM}: mistake
    - Adjusting experimental data \textit{with PSM}: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake
Conclusions

• **The Matching Frontier**
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric

• **Propensity score matching:**
  • The problem:
    • Imbalance can be worse than original data
    • Can increase imbalance when removing the worst matches
    • Approximates random matching in well-balanced data
      (Random matching increases imbalance)
  • Implications:
    • Balance checking required
    • Adjusting for potentially irrelevant covariates *with PSM*: mistake
    • Adjusting experimental data *with PSM*: mistake
    • Reestimating the propensity score after eliminating noncommon support: mistake
    • 1/4 caliper on propensity score: mistake
Conclusions

• **The Matching Frontier**
  • Fast; easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your choice of imbalance metric

• **Propensity score matching:**
  • The problem:
    • Imbalance can be worse than original data
    • Can increase imbalance when removing the worst matches
    • Approximates random matching in well-balanced data
      (Random matching increases imbalance)
  • Implications:
    • Balance checking required
    • Adjusting for potentially irrelevant covariates *with PSM*: mistake
    • Adjusting experimental data *with PSM*: mistake
    • Reestimating the propensity score after eliminating noncommon support: mistake
    • 1/4 caliper on propensity score: mistake

• **Software on its way** · · ·