Simplifying Causal Inference$^1$

Gary King$^2$

Institute for Quantitative Social Science
Harvard University

(Talk at the University of South Carolina, 2/28/2014)

---

$^1$Joint work with Christopher Lucas and Richard Nielsen
$^2$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 \((n-\text{imbalance})\) “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
  - 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

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
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching prunes \( n \) to improve \textit{imbalance}, but
  - Some: set \( n \) and don’t guarantee \textit{imbalance}
  - Others: set \textit{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”
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>Exp</td>
<td>-6.016</td>
<td>3.071</td>
</tr>
<tr>
<td>Decade</td>
<td>-.299</td>
<td>.169</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609</td>
<td>2.157</td>
</tr>
<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>
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)

![Graph showing matching within the interpolation region.](image-url)
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 ($T_i = 1$) 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:
     $$\text{SATT} = \text{mean}_{i \in \{ T_i = 1 \}} (TE_i)$$
  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:
    
    $\text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)$

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

  2. FSATT: Feasible Average Treatment effect on the Treated

  

How Matching Works

- Notation:
  \( Y_i \) Dependent variable

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

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

- Prune unmatched units to improve balance (so \( X_i \) 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)

- **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
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:
    \[ \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved} \]
  - Estimate $Y_i(T_i = 0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls:
    \[ \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \]
  - 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 \( i \):
    \[\text{TE}_i = Y_i (T_i = 1) - Y_i (T_i = 0)\]
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 $i$:
    \[ \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0) \]
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 $i$:
    \[ TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) \]
    = observed $-$ 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
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
    \]
How Matching Works

• Notation:
  \(Y_i\) Dependend 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
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0)
    \]
    (or a model)
  • Prune unmatched units to improve balance (so \(X\) is unimportant)
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
    \[
    \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \text{ (or a model)}
    \]
  - 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
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
    $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (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:
    \[ \text{SATT} = \text{mean}_{i \in \{T_i = 1\}} (TE_i) \]
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
    \[ \hat{Y}_i(T_i = 0) = Y_j(T_i = 0) \]
    (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:

    \[ \text{SATT} = \text{mean}_{i \in \{T_i=1\}} (TE_i) \]

  2. **FSATT**: Feasible Average Treatment effect on the Treated
Method 1: Mahalanobis Distance Matching

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

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($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$
   - 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

Education (years)

Age

Education (years)
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
CC C C C
C
Method 2: Propensity Score Matching

1. Preprocess (Matching)

- Reduce $k$ elements of $X$ to scalar $\pi_i \equiv \Pr(T_i = 1 | X)$
- 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 $> \text{caliper}$

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

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

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

Education (years) vs. Age

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

Graph showing data points with symbols indicating matching criteria.
Propensity Score Matching

Age

Education (years)

Propensity Score
Propensity Score Matching

Age

Education (years)

Propensity Score
Propensity Score Matching

Age

12 16 20 24 ...

Education (years)

Propensity Score

0 1
Propensity Score Matching

Education (years)

Age
12 16 20 24 28
20
30
40
50
60
70
80

Propensity Score

C
CC
C
CC
C
C
CC
C
CC
CC
T
TTT
T
TT
T
T
T
T
TT
T
T
T
T
T
T
T

Education (years)
Propensity Score Matching

Age

12 16 20 24 28

Education (years)

C C C C C C C T T T T T T T

Propensity Score

1 0
Propensity Score Matching

![Propensity Score Matching Graph]

- Education (years) vs. Age
- Symbols indicating data points
- Axes labeled: Age and Education (years)
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

Education vs. Age

Education

Age

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80
Coarsened Exact Matching

Education vs. Age

- Education levels: 12, 14, 16, 18, 20, 22, 24, 26, 28
- Age range: 20 to 80

Graph shows the distribution of education levels across different age groups.
The Matching Frontier

• Bias-Variance trade off
  ⇝ Imbalance-Trade Off

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 $\rightsquigarrow$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
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 ($L_1$):
  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 $n$
- To use, make 3 choices:
  1. Imbalance metric, e.g.:

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 $= \text{matched dataset with lowest imbalance for each } n$
- 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)
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 ($L_1$):

• Simple to use
• All solutions are optimal
• No iteration or diagnostics required
• No cherry picking possible
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

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 ($L_1$):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching

- Simple to use
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible
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:
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 (\(L_1\)):
  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 $\rightarrow$ 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 \(X_0\)

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

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

We develop new algorithms for several frontiers which:
- run very fast
- do not require evaluating every subset
- work with very large data sets

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

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

\(\Rightarrow\) 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

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

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

\(\Rightarrow\) 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!
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!
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!
  - \(\Rightarrow\) 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!
  - \(\implies\) 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

\(\implies\) 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!
  • \(\rightsquigarrow\) 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

Causal effects have big jumps (right)
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

Spline 3
Spline 2
Spline 1

Ethnic Frac.
Cold War
Religious Frac.
Infant Mortality
In GDP p.c.
Mountains
Human Rights
In 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>Initial Effect</th>
<th>Change</th>
<th>New Effect</th>
<th>Remaining Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambia</td>
<td>1991</td>
<td>0.008</td>
<td>0.007</td>
<td>0.015</td>
<td>1608</td>
</tr>
<tr>
<td>Niger</td>
<td>1994</td>
<td>0.015</td>
<td>0.018</td>
<td>0.023</td>
<td>1595</td>
</tr>
<tr>
<td>Lesotho</td>
<td>1998</td>
<td>0.021</td>
<td>0.018</td>
<td>0.018</td>
<td>1254</td>
</tr>
<tr>
<td>Cote D'Ivoire</td>
<td>2002</td>
<td>0.011</td>
<td>0.008</td>
<td>0.008</td>
<td>995</td>
</tr>
<tr>
<td>Guinea</td>
<td>2000</td>
<td>0.005</td>
<td>0.007</td>
<td>0.000</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></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 → 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
### 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

Frontier

Covariate 1

Covariate 2

Treated
Control
Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

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

Number of Observations Dropped
Average Mahalanobis Discrepancy

0
0.1
0.2
0.3
0.4
0.5
1.0

0
0.0
0.1
0.2
0.3
0.4

Field
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
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

Covariate 1
Covariate 2
- Treated
- Control
- Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2
-1.0 -0.5 0.0 0.5 1.0
-1.0
-0.5
0.0
0.5
1.0
●
●
●
●
●
●
●
●
●
●

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

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

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

Remaining Data

Frontier

-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 ●
●
●●
●
●
●
●

Covariate 1
Covariate 2
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

○ Treated
○ Control
○ Next to remove

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

Graph showing the remaining data and the frontier with treated, control, and next to remove observations.
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

●

○

Next to remove

○

Control

Treated

Number of Observations Dropped

Average Mahalanobis Discrepancy

0

0.1

0.2

0.3

0.4

0

5

10

15

20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

○ Treated
○ Control
○ Next to remove

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

○ Treated
□ Control
□ Next to remove

Frontier

- Average Mahalanobis Discrepancy

- Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

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

Remaining Data

Frontier

Treated
Control
Next to remove

Covariate 1
Covariate 2
−1.0 −0.5 0.0 0.5 1.0
−1.0
−0.5
0.0
0.5
1.0

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

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

Frontier

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

Remaining Data

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

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

Graph showing the remaining data distribution and the frontier for observations dropped.
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 and scatter plot showing frequency distribution across different bins with legend for Treatment and Control groups.](chart.png)
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

<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>2</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>

- **Treatment**
- **Control**

Number of Observations Dropped vs L2
Constructing the L1/L2 SATT Frontier

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

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

![Bar chart showing frequency distribution for bins Bin1 to Bin6]  
- Bin1: 4 Treatment, 4 Control
- Bin2: 6 Treatment, 7 Control
- Bin3: 2 Treatment, 3 Control
- Bin4: 3 Treatment, 4 Control
- Bin5: 3 Treatment, 2 Control
- Bin6: 1 Treatment, 2 Control

![Line graph showing L2 penalty vs number of observations dropped]  
- L2 penalty decreases as the number of observations dropped increases.

Legend:  
- Blue: Treatment
- Red: Control
Constructing the L1/L2 SATT Frontier

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

**Graph:**
- **Y-axis:** L2
- **X-axis:** Number of Observations Dropped
- **Points:**
  - L2 values at 0, 0.10, 0.08, 0.06, 0.04, 0.02
  - Number of observations dropped: 0, 2, 4, 6, 8, 10
Constructing the L1/L2 SATT Frontier

Frequency

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

Number of Observations Dropped

L2

Number of Observations Dropped
Constructing the L1/L2 SATT Frontier

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

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin 1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Bin 2</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Bin 3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bin 4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bin 5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Bin 6</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- **L2**: 0.12 / 0.00

- **Number of Observations Dropped**

- **L2** vs. **Number of Observations Dropped**
Constructing the L1/L2 SATT Frontier

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

- **Frequency Distribution**: The bars indicate the number of observations in each bin for both Treatment and Control groups.
- **L2 Values**: The line graph shows the decrease in L2 values as the number of observations dropped increases.

**Legend**:
- Blue bars: Treatment
- Red bars: Control
Constructing the L1/L2 SATT Frontier

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

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

Warning: This figure omits some technical details too!

Works very fast, even with very large data sets
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>$L_1$</th>
<th>Difference in Means</th>
</tr>
</thead>
</table>

- **Published PSM**
- **Published PSM with 1/4 sd caliper**
- **Raw Data**
- **Random Pruning**
- "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>

- **Mahalanobis Discrepancy**
  - Raw Data
  - Random Pruning

- **L₁**
  - "Best Practices" PSM
  - PSM

- **Difference in Means**
  - MDM
  - CEM

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

PSM Matches
CEM and MDM Matches
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:
  - 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···
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 *with PSM*: mistake
    - Adjusting experimental data *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**
For more information

GaryKing.org