Optimizing Balance and Sample Size in Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

(Talk at the Institute for Health Metrics and Evaluation, 6/10/2013)

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 methods optimize only 1 of 2 parameters

Some fix

and don't guarantee imbalance

Others fix imbalance

and don't guarantee n

(Plus: Matching methods optimize a different “imbalance” than post-hoc checks)

Solution: A New Approach — Easier & More Powerful

We estimate the “n-imbalance 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 methods optimize only 1 of 2 parameters

  - Some fix and don't guarantee imbalance
  - Others fix imbalance and don't guarantee

  - (Plus: Matching methods optimize a different "imbalance" than post-hoc checks)

- Solution: A New Approach — Easier & More Powerful

  - We estimate the "$n$-imbalance 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 methods optimize only 1 of 2 parameters

Some fix

Others fix

(Plus: Matching methods optimize a different "imbalance" than post-hoc checks)

Solution: A New Approach — Easier & More Powerful

We estimate the "n-imbalance 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 methods optimize only 1 of 2 parameters
  - Some fix \( n \) and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee \( n \)

- Solution: A New Approach — Easier & More Powerful
  - We estimate the "\( n \)-imbalance 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 methods optimize only 1 of 2 parameters
  - Some fix \( n \) and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee \( n \)
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix $n$ and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee $n$
  - (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix $n$ and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee $n$
  - (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching methods optimize only 1 of 2 parameters
  • Some fix \( n \) and don’t guarantee imbalance
  • Others fix imbalance and don’t guarantee \( n \)
  • (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
• Solution: A New Approach — Easier & More Powerful
  • We estimate the “\( n \)-imbalance frontier”
Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix $n$ and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee $n$
  - (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the “$n$-imbalance frontier”
  - Imbalance metric choice defines the frontier
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching methods optimize only 1 of 2 parameters
  • Some fix $n$ and don’t guarantee imbalance
  • Others fix imbalance and don’t guarantee $n$
  • (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
• Solution: A New Approach — Easier & More Powerful
  • We estimate the “$n$-imbalance frontier”
  • Imbalance metric choice defines the frontier
• Side point:
Overview

• Problem: Model dependence (review)
• Solution: Matching to reduce model dependence (review)
• Problem: Matching methods optimize only 1 of 2 parameters
  • Some fix \( n \) and don’t guarantee imbalance
  • Others fix imbalance and don’t guarantee \( n \)
  • (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
• Solution: A New Approach — Easier & More Powerful
  • We estimate the “\( n \)-imbalance 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 methods optimize only 1 of 2 parameters
  - Some fix $n$ and don’t guarantee imbalance
  - Others fix imbalance and don’t guarantee $n$
  - (Plus: Matching methods optimize a different “imbalance” than post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the “$n$-imbalance 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)
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>-0.445</td>
<td>.126</td>
</tr>
<tr>
<td>Wardur</td>
<td>0.006</td>
<td>.006</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.259</td>
<td>.703</td>
</tr>
<tr>
<td>Factnum2</td>
<td>0.062</td>
<td>.065</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>0.004</td>
<td>.002</td>
</tr>
<tr>
<td>Develop</td>
<td>0.001</td>
<td>.000</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.016</td>
<td>3.071</td>
</tr>
<tr>
<td>Decade</td>
<td>-0.299</td>
<td>.169</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609</td>
<td>2.157</td>
</tr>
</tbody>
</table>

| N              | 122   | 122   |
| Log-likelihood | -45.649 | -44.902 |
| Pseudo $R^2$   | .423  | .433  |
In Sample Fit

Counterfactual Prediction

Doyle and Sambanis: Model Dependence
Model Dependence: A Simpler Example

What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 Political Analysis)
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 Political Analysis)

What to do?
• Preprocess I: Eliminate extrapolation region
• Preprocess II: Match (prune bad matches) within interpolation region
• Model remaining imbalance
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 Political Analysis)

What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance

Dashed: quadratic
Solid: linear (dotted: 95% CI)

Treatment group data
Control group data
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 Political Analysis)

What to do?

- Preprocess I: Eliminate extrapolation region
What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance
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

- 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)

- QoI: Sample Average Treatment effect on the Treated:
  \[ \text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(TE_i) \]

- or Feasible Average Treatment effect on the Treated: FSATT
How Matching Works

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

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

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

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

- **QoI: Sample Average Treatment effect on the Treated**:
  
  $$\text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(TE_i)$$

- **or 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)

- **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)

- **QoI:** Sample Average Treatment effect on the Treated:
  \[ SATT = \text{mean}_{i \in \{T_i = 1\}}(TE_i) \]

- **or Feasible Average Treatment effect on the Treated:**
  $FSATT$
How Matching Works

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

- 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)

- QoI: Sample Average Treatment effect on the Treated:
  \[ SATT = \text{mean}_{i \in \{T_i = 1\}} (TE_i) \]

- or Feasible Average Treatment effect on the Treated: $FSATT$
How Matching Works

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

• 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)

• QoI: Sample Average Treatment effect on the Treated:

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

• or Feasible Average Treatment effect on the Treated: FSATT
How Matching Works

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

• 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

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

  \[
  \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}
  \]
How Matching Works

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

- **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**
  \((X_i \approx X_j)\)

- **QoI**: Sample Average Treatment effect on the Treated:
  \[
  \text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(\text{TE}_i)
  \]

- **Feasible Average Treatment effect on the Treated**: \( \text{FSATT} \)
How Matching Works

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

- **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

• **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)

• QoI: Sample Average Treatment effect on the Treated:

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

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

- **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)

- QoI: Sample Average Treatment effect on the Treated:

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

- or Feasible Average Treatment effect on the Treated: $FSATT$
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\( (X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)
   - 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)
   - $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if $\text{Distance} > \text{caliper}$

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

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

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

12 14 16 18 20 22 24 26 28

20 30 40 50 60 70 80

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C

C
Mahalanobis Distance Matching

Age
12 14 16 18 20 22 24 26 28

Education (years)
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
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 $\pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}$
   - 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 $> \text{caliper}$

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
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-x_i\beta}} \]

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

Age

Education (years)
Propensity Score Matching

Age vs. Education (years)

Propensity Score
Propensity Score Matching

Age

Education (years)

Propensity Score
Propensity Score Matching

Education (years) vs. Age

- Ages range from 12 to 80
- Education levels range from 12 to 28 years
- Propensity scores indicated by 'C' for control and 'T' for treatment
Propensity Score Matching

Education (years) vs. Age

Age

12 16 20 24 28

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)
   - 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

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

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: CoarsenedExact 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 levels: 12, 14, 16, 18, 20, 22, 24, 26, 28
- Age ranges: 20, 30, 40, 50, 60, 70, 80

Graph showing the distribution of individuals across different education levels and age groups.
Coarsened Exact Matching

Education

- HS
- BA
- MA
- PhD
- 2nd PhD

Drinking age

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

C

C

T

T

C

C

T

T

C

C

T

T

C

C
Coarsened Exact Matching

Age

Education

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

C

C

C

CC

C

C

C

C

T

T

T

T
The Matching Frontier

• Bias-Variance trade off
  ⇝ Imbalance-Trade Off
• Choose an imbalance metric
  • Classic measure: Difference of means (for each variable)
  • Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  • Difference of multivariate histograms (L1):
    \[ L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H} |f_{\ell_1} \cdots f_{\ell_k} - g_{\ell_1} \cdots g_{\ell_k}| \]
  • Difference of multivariate histograms (L2)
  • The metric defines the "n-imbalance frontier" (lowest imbalance for each n)
• Choose a matching solution (trading off bias and variance)
• Result: Optimal. No need to iterate. Choice of solution left to researcher.
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
- Choose an imbalance metric

$\text{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1, \ldots, \ell_k \in H}(X) |f_{\ell_1, \ldots, \ell_k} - g_{\ell_1, \ldots, \ell_k}|$

Difference of multivariate histograms (L2)
The Matching Frontier

- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
The Matching Frontier

- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - 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
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms ($L_1$):
    \[
    L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(X)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
    \]
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):
    \[
    \mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(X)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
    \]
  - Difference of multivariate histograms (L2)
The Matching Frontier

• Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off

• Choose an imbalance metric
  • Classic measure: Difference of means (for each variable)
  • Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  • Difference of multivariate histograms (L1):

$$L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(X)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|$$

• Difference of multivariate histograms (L2)

• The metric defines the “$n$-imbalance frontier” (lowest imbalance for each $n$)
The Matching Frontier

- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):
    \[
    L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(\mathbf{x})} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
    \]
  - Difference of multivariate histograms (L2)
- The metric defines the “$n$-imbalance frontier” (lowest imbalance for each $n$)
- Choose a matching solution (trading off bias and variance)
The Matching Frontier

• Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
• Choose an imbalance metric
  • Classic measure: Difference of means (for each variable)
  • Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  • Difference of multivariate histograms (L1):
    \[
    L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(X)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
    \]
  • Difference of multivariate histograms (L2)
• The metric defines the “$n$-imbalance frontier” (lowest imbalance for each $n$)
• Choose a matching solution (trading off bias and variance)
• Result: Optimal. No need to iterate. Choice of solution left to researcher.
Example Frontier, and Results

Frontier

Mahalanobis Discrepancy

N of matched sample

Estimated Treatment Effect

logit coefficient

N of matched sample
Constructing the Mahalanobis Frontier
Constructing the Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0 -0.5 0.0 0.5 1.0

Treated
Control
Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the Mahalanobis Frontier

Remaining Data

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

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

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

Remaining Data

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

Frontier

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

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2

-1.0 -0.5 0.0 0.5 1.0

●
●
●
●
●
●
●

Treated
Control
Next to remove

Frontier

Average Mahalanobis Discrepancy

0.4
0.3
0.2
0.1
0.0

0 5 10 15 20

Number of Observations Dropped

0

0.0

0.1

0.2

0.3

0.4

0.0

0.1

0.2

0.3

0.4
Constructing the Mahalanobis Frontier

**Remaining Data**
- **Covariate 1**
- **Covariate 2**
- -1.0, -0.5, 0.0, 0.5, 1.0

- ●
- ○
- ▲
- ○
- ○
- ○
- ○
- ●
- ●
- ●
- ●
- ●

**Frontier**
- **Average Mahalanobis Discrepancy**
- ○
- ○
- ●
- ●
- ●
- ●
- ●
- ●
- ●
- ●

**Number of Observations Dropped**
- 0, 5, 10, 15, 20

- ○
- ○
- ●
- ●
- ●
- ●
- ●
- ●
- ●
- ●

- ○
- ○
- ●
- ●
- ●
- ●
- ●
- ●
- ●
- ●
Constructing the Mahalanobis Frontier

**Remaining Data**

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

**Frontier**

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

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the Mahalanobis Frontier

Remaining Data

Frontier

- Treating
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

Covariate 1

Covariate 2
Constructing the 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

0.4

0.3

0.2

0.1

0.0

0  5  10  15  20

Number of Observations Dropped

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

●

●

●

●

●

●

●

●

●

●

●
Constructing the Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1

Covariate 2

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the 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

0.0

0.1

0.2

0.3

0.4

0.0

0.1

0.2

0.3

0.4

0.0

0.5

1.0

Number of Observations Dropped

Average Mahalanobis Discrepancy

0 5 10 15 20
Constructing the Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1
Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0
-0.5
0.0
0.5
1.0

● ●
●
●
●
●
●
●
●

Treated
Control
Next to remove

Number of Observations Dropped
Average Mahalanobis Discrepancy

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the Mahalanobis Frontier

Remaining Data

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

Next to remove

Frontier

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

Legend:
- ○ Treated
- ○ Control
- ○ Next to remove
Constructing the L1/L2 Frontier

![Graph showing frequency and L2 values across different bins]

- **Bins**: Bin 1 to Bin 6
- **Frequency Counts**:
  - 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
- **L2 Values**:
  - Bin 1: 0.00
  - Bin 2: 0.00
  - Bin 3: 0.00
  - Bin 4: 0.00
  - Bin 5: 0.00
  - Bin 6: 0.00

Note: The figure includes a bar chart and a scatter plot with L2 values on a logarithmic scale.
Constructing the L1/L2 Frontier

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

- **Number of Observations Dropped**
  - L2: 38 / 56
Constructing the L1/L2 Frontier

![Graph showing frequency of observations in different bins for Treatment and Control groups. The x-axis represents bins labeled Bin1 to Bin6, and the y-axis represents frequency. The Treatment group is represented by blue bars, while the Control group is represented by red bars. The graph also includes a scatter plot showing the number of observations dropped against L2 value, with a downward trend.]
Constructing the L1/L2 Frontier
Constructing the L1/L2 Frontier

Frequency

<table>
<thead>
<tr>
<th>Bin</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>4</td>
</tr>
<tr>
<td>Bin2</td>
<td>6</td>
</tr>
<tr>
<td>Bin3</td>
<td>7</td>
</tr>
<tr>
<td>Bin4</td>
<td>3</td>
</tr>
<tr>
<td>Bin5</td>
<td>3</td>
</tr>
<tr>
<td>Bin6</td>
<td>2</td>
</tr>
</tbody>
</table>

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

Log-log graph showing the number of observations dropped over L2, with a decreasing trend from 0.12 to 0.00 as the number of observations dropped increases from 0 to 10.
Constructing the L1/L2 Frontier

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

- **Graph**: 
  - X-axis: Number of Observations Dropped
  - Y-axis: L2
  - Data points: (0, 0.12), (2, 0.10), (4, 0.08), (6, 0.06), (8, 0.04), (10, 0.02)
Constructing the L1/L2 Frontier

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

- **Number of Observations Dropped**: L2: 43 / 56
- **Axis Labels**:
  - Frequency
  - L2
  - Number of Observations Dropped

- **Legend**:
  - Treatment
  - Control
## Constructing the L1/L2 Frontier

<table>
<thead>
<tr>
<th>Bin</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>4</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>

### Frequency Distribution

- **Bin1**: Treatment = 4, Control = 4
- **Bin2**: Treatment = 6, Control = 6
- **Bin3**: Treatment = 2, Control = 2
- **Bin4**: Treatment = 3, Control = 3
- **Bin5**: Treatment = 3, Control = 3
- **Bin6**: Treatment = 1, Control = 2

### L2 vs. Number of Observations Dropped

- **Number of Observations Dropped**:
  - 0
  - 2
  - 4
  - 6
  - 8
  - 10

- **L2 Values**:
  - 0.00
  - 0.02
  - 0.04
  - 0.06
  - 0.08
  - 0.10
  - 0.12

---

[Graph showing the frequency distribution and the L2 vs. number of observations dropped]
Constructing the L1/L2 Frontier

![Graph showing frequency and L2 values in different bins]

- Bin 1: Frequency (4), L2: 45/56
- Bin 2: Frequency (6), L2: 45/56
- Bin 3: Frequency (2), L2: 45/56
- Bin 4: Frequency (3), L2: 45/56
- Bin 5: Frequency (2), L2: 45/56
- Bin 6: Frequency (1), L2: 45/56

Legend:
- Treatment
- Control
Constructing the L1/L2 Frontier
Foreign Aid Shocks & Conflict
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>

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

Published PSM with 1/4 sd caliper

N of Matched Sample

2500 1500 500 0

0 0.10 0.20 0.30
Healthways Data
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>
<tbody>
<tr>
<td>N of Matched Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>8</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Raw Data
- Random Pruning
- "Best Practices" PSM
- MDM
- PSM
- CEM
Imbalance Metric

Mahalanobis Discrepancy  |  $L_1$  |  Difference in Means

*Called/Not Called Data*

King, Nielsen, Coberley, Pope, and Wells (2012)
# FDA Drug Approval Times

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>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph 1" /></td>
<td><img src="image2.png" alt="Graph 2" /></td>
<td><img src="image3.png" alt="Graph 3" /></td>
</tr>
</tbody>
</table>

- **Raw Data**
- **Random Pruning**
- **"Best Practices" PSM**
- **MDM**
- **PSM**
- **CEM**
Job Training (Lelonde 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>
<tbody>
<tr>
<td>N of Matched Sample</td>
<td></td>
<td>N of Matched Sample</td>
</tr>
</tbody>
</table>

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

N of Matched Sample
"Best Practices" PSM
PSM
MDM
CEM
PSM Approximates Random Matching in Balanced Data

![Diagram showing PSM Matches and CEM and MDM Matches]
Destroying CEM with PSM’s Two Step Approach

![Diagram showing CEM Matches vs. CEM-generated PSM Matches](image-url)
Conclusions

• The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
      - 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
Conclusions

• The Matching Frontier
  • Easy to use; no need to iterate
Conclusions

• The Matching Frontier
  • Easy to use; no need to iterate
  • No need to choose among matching methods

• Software on its way
Conclusions

- **The Matching Frontier**
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
Conclusions

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

- Propensity score matching:
  - The problem:
Conclusions

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

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

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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**
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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**
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
  • Easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your 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**
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
  • Easy to use; no need to iterate
  • No need to choose among matching methods
  • Optimal results for your 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**
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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
Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your 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