Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

(Talk at University of Rochester, 11/4/2011)
Problem: Model dependence (review)

Solution: Matching to preprocess data (review)

Problem: Many matching methods & specifications

Solution: The Space Graph helps us choose

Problem: The most commonly used method can increase imbalance!

Solution: Other methods do not share this problem

(Coarsened Exact Matching is simple, easy, and powerful)

⇝

Lots of insights revealed in the process
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Many matching methods & specifications
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Many matching methods & specifications
Solution: The Space Graph helps us choose
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Many matching methods & specifications
Solution: The Space Graph helps us choose
Problem: The most commonly used method can increase imbalance!
Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Many matching methods & specifications
Solution: The Space Graph helps us choose
Problem: The most commonly used method can increase imbalance!
Solution: Other methods do not share this problem
(Coarsened Exact Matching is simple, easy, and powerful)
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Problem: Many matching methods & specifications
Solution: The Space Graph helps us choose
Problem: The most commonly used method can increase imbalance!
Solution: Other methods do not share this problem
(Coarsened Exact Matching is simple, easy, and powerful)
⇒ Lots of insights revealed in the process
Model Dependence Example

Data:
- 124 Post-World War II civil wars

Dependent variable:
- Peacebuilding success

Treatment variable:
- Multilateral UN peacekeeping intervention (0/1)

Control vars:
- War type, severity, duration
- Development status, etc.

Counterfactual question:
- UN intervention switched for each war

Data analysis:
- Logit model

The question:
- How model dependent are the results?
Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000

Data:
124 Post-World War II civil wars

Dependent variable:
peacebuilding success

Treatment variable:
multilateral UN peacekeeping intervention (0/1)

Control vars:
war type, severity, duration; development status; etc.

Counterfactual question:
UN intervention switched for each war

Data analysis:
Logit model

The question:
How model dependent are the results?
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data:** 124 Post-World War II civil wars
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data:** 124 Post-World War II civil wars
- **Dependent variable:** peacebuilding success
Data: 124 Post-World War II civil wars
Dependent variable: peacebuilding success
Treatment variable: multilateral UN peacekeeping intervention (0/1)
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data**: 124 Post-World War II civil wars
- **Dependent variable**: peacebuilding success
- **Treatment variable**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status; etc.

Counterfactual question: UN intervention switched for each war

Data analysis: Logit model

The question: How model dependent are the results?
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data:** 124 Post-World War II civil wars
- **Dependent variable:** peacebuilding success
- **Treatment variable:** multilateral UN peacekeeping intervention (0/1)
- **Control vars:** war type, severity, duration; development status; etc.
- **Counterfactual question:** UN intervention switched for each war
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data**: 124 Post-World War II civil wars
- **Dependent variable**: peacebuilding success
- **Treatment variable**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status; etc.
- **Counterfactual question**: UN intervention switched for each war
- **Data analysis**: Logit model
Model Dependence Example
Replication: Doyle and Sambanis, APSR 2000

- **Data**: 124 Post-World War II civil wars
- **Dependent variable**: peacebuilding success
- **Treatment variable**: multilateral UN peacekeeping intervention (0/1)
- **Control vars**: war type, severity, duration; development status; etc.
- **Counterfactual question**: UN intervention switched 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>
<tr>
<td>N</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-45.649</td>
<td></td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.423</td>
<td></td>
</tr>
</tbody>
</table>
Doyle and Sambanis: Model Dependence

In Sample Fit

Counterfactual Prediction

Probabilities from original model vs. probabilities from modified model.
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)

What to do?

Preprocess I: Eliminate extrapolation region

Preprocess II: Match (prune bad matches) within interpolation region

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

Model remaining imbalance
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)

**Graph:**
- **Outcome** on the y-axis
- **Education (years)** on the x-axis
- Data points are marked with 'T'.

**Table:**

<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>T</td>
</tr>
<tr>
<td>14</td>
<td>T</td>
</tr>
<tr>
<td>16</td>
<td>T</td>
</tr>
<tr>
<td>18</td>
<td>T</td>
</tr>
<tr>
<td>20</td>
<td>T</td>
</tr>
<tr>
<td>22</td>
<td>T</td>
</tr>
<tr>
<td>24</td>
<td>T</td>
</tr>
<tr>
<td>26</td>
<td>T</td>
</tr>
<tr>
<td>28</td>
<td>T</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*)

```
<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>
```

Gary King (Harvard, IQSS)
Matching reduces model dependence, bias, and variance
What Matching Does

Matching Notation:
- $Y_i$: Dependent variable
- $T_i$: Treatment variable (0/1)
- $X_i$: Pre-treatment covariates

Treatment Effect for treated ($T_i = 1$) observation:
$$TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

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

or a model:
$$\hat{Y}_i(0) = \hat{g}_0(X_j)$$

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

QoI: Sample Average Treatment effect on the Treated:
$$SATT = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i$$

or Feasible Average Treatment effect on the Treated:
$$FSATT$$

Gary King (Harvard, IQSS)
Matching Methods
What Matching Does

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

Treatment Effect for treated ($T_i = 1$) observation:

$$TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}$$

Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls

$$\hat{Y}_i(0) = Y_j(0)$$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

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

QoI: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i$$

or Feasible Average Treatment effect on the Treated: $\text{FSATT}$
What Matching Does

- **Notation:**
  - $Y_i$: Dependent variable
What Matching Does

- Notation:
  - \( Y_i \): Dependent variable
  - \( T_i \): Treatment variable (0/1)

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 \( Y_i(0) \) with \( Y_j(0) \) from matched (\( X_i \approx X_j \)) controls:
\[
\hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j)
\]

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

QoI: Sample Average Treatment effect on the Treated:
\[
\text{SATT} = \frac{1}{n_T} \sum_{i \in \{ T_i = 1 \}} TE_i
\]
or Feasible Average Treatment effect on the Treated:
\[
\text{FSATT}
\]
What Matching Does

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

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 $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls:

$\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

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

QoI: Sample Average Treatment effect on the Treated:

$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i$

or Feasible Average Treatment effect on the Treated: $\text{FSATT}$
What Matching Does

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

- Treatment Effect for treated ($T_i = 1$) observation $i$:

\[
\hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j)
\]

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

QoI: Sample Average Treatment effect on the Treated:

\[
\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i
\]

or Feasible Average Treatment effect on the Treated:
What Matching Does

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

- Treatment Effect for treated \((T_i = 1)\) observation \(i\):
  \[
  \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)
  \]
What Matching Does

- Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1)
  - $X_i$: 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 $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls

$\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

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

QoI: Sample Average Treatment effect on the Treated:

$$SATT = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i$$

or Feasible Average Treatment effect on the Treated: $FSATT$
What Matching Does

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

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

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
  \[
  \hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j)
  \]

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

QoI: Sample Average Treatment effect on the Treated: $\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i$

or Feasible Average Treatment effect on the Treated: $\text{FSATT}$
What Matching Does

- **Notation:**
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1)
  - $X_i$: 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 $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls**
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

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

- Notation:
  - $Y_i$: Dependent variable
  - $T_i$: Treatment variable (0/1)
  - $X_i$: 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 $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls:
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

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

- QoI: Sample Average Treatment effect on the Treated:
  \[
  \text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i
  \]
What Matching Does

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

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

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
  \[ \hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j) \]

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

- QoI: Sample Average Treatment effect on the Treated:
  \[ SATT = \frac{1}{n_T} \sum_{i \in \{T_i=1\}} TE_i \]

- or Feasible Average Treatment effect on the Treated: $FSATT$
Method 1: Mahalanobis Distance Matching

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
Method 1: Mahalanobis Distance Matching

1. **Preprocess** (Matching)

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

Gary King (Harvard, IQSS)
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
Mahalanobis Distance Matching
Mahalanobis Distance Matching

Education (years)

Age

12 14 16 18 20 22 24 26 28

20
30
40
50
60
70
80

Gary King (Harvard, IQSS)

Matching Methods
Mahalanobis Distance Matching

Education (years)

Age

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

21 / 57
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

Preprocess

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
Method 2: Propensity Score Matching

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
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
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
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
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
**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
Propensity Score Matching

Age

Education (years)

Propensity Score

Gary King (Harvard, IQSS)

Matching Methods
Propensity Score Matching

Education (years)

Age

12 16 20 24 28

20

30

40

50

60

70

80

0

1

Propensity Score

Gary King (Harvard, IQSS)

Matching Methods
Propensity Score Matching

Age

Education (years)

Propensity Score

Gary King (Harvard, IQSS)

Matching Methods
Propensity Score Matching

Gary King (Harvard, IQSS)

Matching Methods
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)
Method 3: Coarsened Exact Matching

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
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
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
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
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
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
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
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
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
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)
Coarsened Exact Matching
Coarsened Exact Matching
Coarsened Exact Matching

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

Education
HS BA MA PhD 2nd PhD
Drinking age

Gary King (Harvard, IQSS)
Matching Methods
Coarsened Exact Matching

Education

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

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

Gary King (Harvard, IQSS)
None
The Bias-Variance Trade Off in Matching

Bias (model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)

\( \Rightarrow \) we measure imbalance instead

Variance = \( f(\text{matched sample size}, \text{estimator}) \)

\( \Rightarrow \) we measure matched sample size instead

Bias-Variance trade off = Imbalance-Trade Off

Measuring Imbalance

Classic measure: Difference of means (for each variable)

Better measure (difference of multivariate histograms):

\[ L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1, \ldots, \ell_k \in H} |f_{\ell_1, \ldots, \ell_k}(X) - g_{\ell_1, \ldots, \ell_k}(X)| \]
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)
  \[ \Rightarrow \text{we measure } \text{imbalance} \text{ instead} \]
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)
  \( \Rightarrow \) we measure *imbalance* instead

- **Variance** = \( f(\text{matched sample size}, \text{estimator}) \)
  \( \Rightarrow \) we measure *matched sample size* instead

---

Gary King (Harvard, IQSS)
The Bias-Variance Trade Off in Matching

- **Bias** (\& model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)
  \( \rightsquigarrow \) we measure \text{imbalance} instead

- **Variance** = \( f(\text{matched sample size}, \text{estimator}) \)
  \( \rightsquigarrow \) we measure \text{matched sample size} instead

- **Bias-Variance trade off** \( \rightsquigarrow \) **Imbalance-\( n \) Trade Off**
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = $f(\text{imbalance, importance, estimator})$
  $\sim$ we measure *imbalance* instead
- **Variance** = $f(\text{matched sample size, estimator})$
  $\sim$ we measure *matched sample size* instead
- **Bias-Variance trade off** $\sim$ **Imbalance-$n$ Trade Off**
- Measuring Imbalance
Bias (model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)
\[ \implies \text{we measure imbalance instead} \]

Variance = \( f(\text{matched sample size}, \text{estimator}) \)
\[ \implies \text{we measure matched sample size instead} \]

Bias-Variance trade off \( \implies \) Imbalance-\( n \) Trade Off

Measuring Imbalance
- Classic measure: Difference of means (for each variable)
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)
  \( \leadsto \) we measure **imbalance** instead

- **Variance** = \( f(\text{matched sample size}, \text{estimator}) \)
  \( \leadsto \) we measure **matched sample size** instead

- **Bias-Variance trade off** \( \leadsto \) **Imbalance-\( n \)** Trade Off

- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)
  - Better measure (difference of multivariate histograms):

\[
\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(X)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
\]
Comparing Matching Methods

MDM & PSM: Choose matched $n$, match, check imbalance

CEM: Choose imbalance, match, check matched $n$

Best practice: iterate
Choose matched solution & matching method becomes irrelevant

Our idea: Compute lots of matching solutions, identify the frontier of lowest imbalance for each given $n$, and choose a matching solution
MDM & PSM: Choose matched $n$, match, check imbalance

Best practice: iterate Choose matched solution & matching method becomes irrelevant

Our idea: Compute lots of matching solutions, identify the frontier of lowest imbalance for each given $n$, and choose a matching solution
Comparing Matching Methods

- MDM & PSM: Choose matched $n$, match, check imbalance
- CEM: Choose imbalance, match, check matched $n$
Comparing Matching Methods

- MDM & PSM: Choose matched $n$, match, check imbalance
- CEM: Choose imbalance, match, check matched $n$
- Best practice: iterate
Comparing Matching Methods

- MDM & PSM: Choose matched $n$, match, check imbalance
- CEM: Choose imbalance, match, check matched $n$
- Best practice: iterate
- Choose matched solution & matching method becomes irrelevant
Comparing Matching Methods

- MDM & PSM: Choose matched $n$, match, check imbalance
- CEM: Choose imbalance, match, check matched $n$
- Best practice: iterate
- Choose matched solution & matching method becomes irrelevant
- Our idea: Compute lots of matching solutions, identify the frontier of lowest imbalance for each given $n$, and choose a matching solution
A Space Graph: Real Data
King, Nielsen, Coberley, Pope, and Wells (2011)

Healthways Data

- Raw Data
- Random Pruning
- PSM
- MDM
- CEM
A Space Graph: Real Data

Lalonde Data Subset

Gary King (Harvard, IQSS)
Space Graphs: Different Imbalance Metrics

Aid Shocks (L1 Metric)

Aid Shocks (Difference in Means Metric)

Aid Shocks (Average Mahalanobis Discrepancy)

Gary King (Harvard, IQSS)
Matching Methods
A Space Graph: Simulated Data — Mahalanobis

MDM: 1 Covariate

- Imbalance:
  - High
  - Med
  - Low

N of matched sample
L1
500 250 0
0.0 0.5 1.0
High
Med
Low

MDM: 2 Covariates

- Imbalance:
  - High
  - Med
  - Low

N of matched sample
L1
500 250 0
0.0 0.5 1.0
High
Med
Low

MDM: 3 Covariates

- Imbalance:
  - High
  - Med
  - Low

N of matched sample
L1
500 250 0
0.0 0.5 1.0
High
Med
Low
A Space Graph: Simulated Data — CEM

CEM: 1 Covariate

CEM: 2 Covariates

CEM: 3 Covariates

Gary King (Harvard, IQSS)
A Space Graph: Simulated Data — Propensity Score

PSM: 1 Covariate

PSM: 2 Covariates

PSM: 3 Covariates

Imbalance:

High
Med
Low
PSM Approximates Random Matching in Balanced Data

Gary King (Harvard, IQSS)

Matching Methods
CEM Weights and Nonparametric Propensity Score

CEM Weight:

\[ w_i = \frac{m_i^T}{m_i^C} \] ( + normalization)

CEM: Gives a better pscore than PSM

Doesn't match based on crippled information
CEM Weights and Nonparametric Propensity Score

CEM Weight:  
\[ w_i = \frac{m_i^T}{m_i^C} \quad (+ \text{normalization}) \]

CEM Pscore:  
\[ \hat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C} \]
CEM Weights and Nonparametric Propensity Score

CEM Weight: \[ w_i = \frac{m_i^T}{m_i^C} \quad (\text{+ normalization}) \]

CEM Pscore: \[ \hat{Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C} \]

\[ \rightsquigarrow \text{CEM:} \]
CEM Weights and Nonparametric Propensity Score

CEM Weight: \( w_i = \frac{m_i^T}{m_i^C} \) (with normalization)

CEM Pscore: \( \hat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C} \)

\( \rightsquigarrow \) CEM:
- Gives a better pscore than PSM
CEM Weights and Nonparametric Propensity Score

CEM Weight: \( w_i = \frac{m_i^T}{m_i^C} \) ( + normalization)

CEM Pscore: \( \hat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C} \)

~ CEM:
- Gives a better pscore than PSM
- Doesn’t match based on crippled information
Destroying CEM with PSM’s Two Step Approach

- CEM Matches
- CEM-generated PSM Matches
Data where PSM Works Reasonably Well — PSM & MDM

Unmatched Data: $L_1 = 0.685$

PSM: $L_1 = 0.452$

MDM: $L_1 = 0.448$

Gary King (Harvard, IQSS)
Matching Methods
Data where PSM Works Reasonably Well — CEM

Bad CEM: L1 = 0.661

Better CEM: L1 = 0.188

Even Better CEM: L1 = 0.095

Gary King (Harvard, IQSS)
Conclusions

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)

The Cause: unnecessary 1st stage dimension reduction

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

In four data sets and many simulations:
- CEM > Mahalanobis > Propensity Score
- (Your performance may vary)

CEM and Mahalanobis do not have PSM's problems
You can easily check with the Space Graph
Conclusions

- 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
    
    \[
    \text{Random matching increases imbalance}
    \]

  - The Cause: unnecessary 1st stage dimension reduction

  - Implications:
    
    \[
    \begin{align*}
    \text{Balance checking required} & \quad \text{Adjusting for potentially irrelevant covariates with PSM: mistake} \\
    & \quad \text{Adjusting experimental data with PSM: mistake} \\
    & \quad \text{Reestimating the propensity score after eliminating noncommon support: mistake} \\
    & \quad \text{1/4 caliper on propensity score: mistake}
    \end{align*}
    \]

  - In four data sets and many simulations: \(\text{CEM} > \text{Mahalanobis} > \text{Propensity Score}\)
    
    \[
    \text{Your performance may vary}
    \]

  - CEM and Mahalanobis do not have PSM's problems
    
    \[
    \text{You can easily check with the Space Graph}
    \]
Conclusions

- Propensity score matching:
  - The problem:
Conclusions

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

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

CEM > Mahalanobis > Propensity Score

CEM and Mahalanobis do not have PSM's problems
You can easily check with the Space Graph
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

- **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)
  - The Cause: unnecessary 1st stage dimension reduction

In four data sets and many simulations:
- CEM \(>\) Mahalanobis \(>\) Propensity Score

(Your performance may vary)

CEM and Mahalanobis do not have PSM's problems
You can easily check with the Space Graph

Gary King (Harvard, IQSS)
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - 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

In four data sets and many simulations:
- CEM > Mahalanobis > Propensity Score (Your performance may vary)

CEM and Mahalanobis do not have PSM’s problems
You can easily check with the Space Graph
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM:* mistake
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
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)
- The Cause: unnecessary 1st stage dimension reduction
- 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

- **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)
  - **The Cause:** unnecessary 1st stage dimension reduction
  - **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

In four data sets and many simulations: CEM > Mahalanobis > Propensity Score

(Your performance may vary)

CEM and Mahalanobis do not have PSM's problems
You can easily check with the Space Graph
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - 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

- In four data sets and many simulations:
  **CEM > Mahalanobis > Propensity Score**
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - 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

- In four data sets and many simulations:
  - **CEM > Mahalanobis > Propensity Score**
  - *(Your performance may vary)*
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - 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

- In four data sets and many simulations:
  - CEM > Mahalanobis > Propensity Score
  - (Your performance may vary)
  - CEM and Mahalanobis do not have PSM’s problems
Conclusions

- **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)
  - The Cause: unnecessary 1st stage dimension reduction
  - 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

- In four data sets and many simulations:
  - CEM > Mahalanobis > Propensity Score
  - *(Your performance may vary)*
  - CEM and Mahalanobis do not have PSM’s problems
  - You can easily check with the Space Graph
For papers, software (for R, Stata, & SPSS), tutorials, etc.

http://GKing.Harvard.edu/cem