Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

(Talk at University of Kansas, 12/2/2011)
Problem: Model dependence (review)
Problem: Model dependence (review)
Solution: Matching to preprocess data (review)
Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications

Gary King (Harvard, IQSS)
Matching Methods
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
(Conasened 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?

Gary King (Harvard, IQSS)
Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
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)
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.
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?
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></th>
<th>Modified Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>SE</td>
<td>P-val</td>
<td>Coeff</td>
</tr>
<tr>
<td>Wartype</td>
<td>-1.742</td>
<td>.609</td>
<td>.004</td>
<td>-1.666</td>
</tr>
<tr>
<td>Logdead</td>
<td>- .445</td>
<td>.126</td>
<td>.000</td>
<td>- .437</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
<td>.258</td>
<td>.006</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.259</td>
<td>.703</td>
<td>.073</td>
<td>-1.045</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.062</td>
<td>.065</td>
<td>.346</td>
<td>.032</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
<td>.010</td>
<td>.004</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
<td>.065</td>
<td>.001</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.016</td>
<td>3.071</td>
<td>.050</td>
<td>-6.215</td>
</tr>
<tr>
<td>Decade</td>
<td>- .299</td>
<td>.169</td>
<td>.077</td>
<td>- .284</td>
</tr>
<tr>
<td>Treaty</td>
<td>2.124</td>
<td>.821</td>
<td>.010</td>
<td>2.126</td>
</tr>
<tr>
<td>UNOP4</td>
<td>3.135</td>
<td>1.091</td>
<td>.004</td>
<td>.262</td>
</tr>
<tr>
<td>Wardur*UNOP4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.037</td>
</tr>
<tr>
<td>Constant</td>
<td>8.609</td>
<td>2.157</td>
<td>.000</td>
<td>7.978</td>
</tr>
<tr>
<td>N</td>
<td>122</td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-45.649</td>
<td></td>
<td></td>
<td>-44.902</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.423</td>
<td></td>
<td></td>
<td>.433</td>
</tr>
</tbody>
</table>
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
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

Gary King (Harvard, IQSS)
What to do?

- Preprocess I: Eliminate extrapolation region
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
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*)

Education (years)

Outcome

Education (years)
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:

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

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

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

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: $\text{FSATT}$
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 \( \hat{Y}_i (T_i = 0) = \hat{g}_0 (X_j) \)

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
How Matching Works

- Notation:
  \( Y_i \): Dependent variable

\[ \text{Treatment Effect for treated (} T_i = 1) \text{ observation} \]
\[ \hat{Y}_i (T_i = 0) = Y_j (T_i = 0) \text{ or a model} \]
\[ \hat{Y}_i (T_i = 0) = \hat{g}_0 (X_j) \]

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

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

or Feasible Average Treatment effect on the Treated:
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:

$$TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

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

$$\hat{Y}_i(T_i = 0) = \hat{g}_0(X_j)$$

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: $FSATT$
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:

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

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

$$\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$$

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

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:

$$FSATT$$
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 $i$:**
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 $i$:

$$TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

  - $= \text{observed} - \text{unobserved}$
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 \[
  \hat{Y}_i(T_i = 0) = \hat{g}_0(X_j)
  \]
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 $i$:**
  \[
  TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)
  \]
  \[
  = \text{observed} - \text{unobserved}
  \]

- **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 \[
  \hat{Y}_i(T_i = 0) = \hat{g}_0(X_j)
  \]

- **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 $i$:
  \[
  \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}
  \]

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

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

- QoI: Sample Average Treatment effect on the Treated:
  \[
  \text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i=1\}} \text{TE}_i
  \]
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 $i$:**
  
  $$TE_i = Y_i(T_i = 1) - Y_i(T_i = 0) = \text{observed} - \text{unobserved}$$

- **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 $\hat{Y}_i(T_i = 0) = \hat{g}_0(X_j)$

- **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:** $\text{FSATT}$
Method 1: Mahalanobis Distance Matching

1. Preprocess

\[ \text{Distance}\left( X^i, X^j \right) = \sqrt{\left( X^i - X^j \right)' S^{-1} \left( X^i - X^j \right)} \]

2. Match each treated unit to the nearest control unit

Control units: not reused; pruned if unused

3. 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
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

Education (years)

Age

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

Education (years)
Mahalanobis Distance Matching

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

Education (years)
Mahalanobis Distance Matching

Education (years)

Age

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

T T

T T

TT TT T TT

TTT TT

T

T TT

C

C

C C

CC

C

C

C CCC

C

C

CC C C C

C

Gary King (Harvard, IQSS)
Method 2: Propensity Score Matching

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
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 $>$ caliper

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

![Graph showing education and age data with match symbols: C for control, T for treatment]
Propensity Score Matching

Education (years)

Age

Propensity Score

Gary King (Harvard, IQSS)

Matching Methods
Propensity Score Matching

Propensity Score

Age

Education (years)

Gary King (Harvard, IQSS)
Propensity Score Matching

Gary King (Harvard, IQSS)
Propensity Score Matching

Age

Education (years)

Propensity Score

Gary King (Harvard, IQSS)

Matching Methods
Method 3: Coarsened Exact Matching

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

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

Education

HS  BA  MA  PhD  2nd PhD

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

Education

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

Drinking age

Education

HS BA MA PhD 2nd PhD

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

Education

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

HS	BA	MA	PhD	2nd PhD

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

Gary King (Harvard, IQSS)

Matching Methods
Coarsened Exact Matching
The Bias-Variance Trade Off in Matching

Bias

\[ \text{Bias} = f(\text{imbalance}, \text{importance}, \text{estimator}) \]

\( \Rightarrow \) we measure imbalance instead

Variance

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

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

Bias-Variance trade off

\( \Rightarrow \) Imbalance-Variance trade off

Measuring Imbalance

Classic measure: Difference of means (for each variable)

Better measure (difference of multivariate histograms):

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

Another measure: Mahalanobis distance to closest unit in other group, averaged over each unit
The Bias-Variance Trade Off in Matching

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

Gary King (Harvard, IQSS)

Matching Methods
The Bias-Variance Trade Off in Matching

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

- **Variance** = \( f(\text{matched sample size}, \text{estimator}) \)
  - \( \rightsarrow \) we measure **matched sample size** instead
The Bias-Variance Trade Off in Matching

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

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

- **Bias-Variance trade off** $\rightsquigarrow$ **Imbalance-\textit{n} Trade Off**
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = $f($imbalance, importance, estimator$)$
  - $\rightsquigarrow$ we measure imbalance instead
- **Variance** = $f($matched sample size, estimator$)$
  - $\rightsquigarrow$ we measure matched sample size instead
- **Bias-Variance trade off** $\rightsquigarrow$ **Imbalance-$n$ Trade Off**
- Measuring Imbalance

- Classic measure: Difference of means (for each variable)
- Better measure (difference of multivariate histograms):
  - $L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$
- Another measure: Mahalanobis distance to closest unit in other group, averaged over each unit

Gary King (Harvard, IQSS)
Bias (\& model dependence) = f(imbalance, importance, estimator) 
\[\rightsquigarrow\text{we measure imbalance instead}\]

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

Bias-Variance trade off \[\rightsquigarrow\text{Imbalance-}n\text{ Trade Off}\]

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

  - Another measure: Mahalanobis distance to closest unit in other group, averaged over each unit
The Bias-Variance Trade Off in Matching

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

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

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

- **Measuring Imbalance**
  - Classic measure: Difference of means (for each variable)
  - Better measure (difference of multivariate histograms):
    \[
    L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \ldots \ell_k \in H(x)} |f_{\ell_1 \ldots \ell_k} - g_{\ell_1 \ldots \ell_k}|
    \]

  - Another measure: Mahalanobis distance to closest unit in other group, averaged over each unit
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.
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

N of Matched Sample ("variance")
L1 ("bias")

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

Called/Not Called Data

N of Matched Sample ("variance")

L1 ("bias")

Gary King (Harvard, IQSS)
A Space Graph: Real Data

Lalonde Data Subset

N of Matched Sample ("variance")

L1 ("bias")

Gary King (Harvard, IQSS)

Matching Methods
Space Graphs: Different Imbalance Metrics

Aid Shocks (L1 Metric)

Aid Shocks (Difference in Means Metric)

Aid Shocks (Average Mahalanobis Discrepancy)
A Space Graph: Simulated Data — Mahalanobis

MDM: 1 Covariate

MDM: 2 Covariates

MDM: 3 Covariates

Imbalance:
- High
- Med
- Low

Matching Methods
CEM: 1 Covariate
N of matched sample
L1
0.0 0.5 1.0
High
Med
Low
Imbalance:

CEM: 2 Covariates
N of matched sample
L1
0.0 0.5 1.0
High
Med
Low
Imbalance:

CEM: 3 Covariates
N of matched sample
L1
0.0 0.5 1.0
High
Med
Low
Imbalance:
A Space Graph: Simulated Data — Propensity Score

Gary King (Harvard, IQSS)  Matching Methods
PSM Approximates Random Matching in Balanced Data

![Graph showing PSM Matches and CEM and MDM Matches]

- PSM Matches
- CEM and MDM Matches
CEM Weight: \[ w_i = \frac{m_i^T}{m_i^C} \] ( + normalization)
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 Weights and Nonparametric Propensity Score

CEM Weight: \( w_i = \frac{m_{iT}}{m_{iC}} \) (plus normalization)

CEM Pscore: \( \hat{Pr}(T_i = 1|X_i) = \frac{m_{iT}}{m_{iT} + m_{iC}} \)

\( \leadsto \) CEM:
CEM Weights and Nonparametric Propensity Score

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

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

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

\[ \sim \text{CEM:} \]
- Gives a better pscore than PSM
- Doesn’t match based on crippled information
Destroying CEM with PSM’s Two Step Approach

![Diagram showing the relationship between Covariate 1 and Covariate 2, with matches indicated by black and red dots. The red dots represent CEM Matches, and the black dots represent CEM-generated PSM Matches.](image-url)
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:

  - 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)

  - 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:
**Conclusions**

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

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
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

- **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
(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
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

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
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
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 \textit{with PSM}: mistake
    - Adjusting experimental data \textit{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:
  \textbf{CEM} > \textbf{Mahalanobis} > \textbf{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
Data where PSM Works Reasonably Well — PSM & MDM

Unmatched Data: $L_1 = 0.685$

PSM: $L_1 = 0.452$

MDM: $L_1 = 0.448$
Data where PSM Works Reasonably Well — CEM

Bad CEM: \( L_1 = 0.661 \)

Better CEM: \( L_1 = 0.188 \)

Even Better CEM: \( L_1 = 0.095 \)