Comparative Effectiveness of Matching Methods for Causal Inference

Gary King
Institute for Quantitative Social Science
Harvard University

joint work with
Richard Nielsen (Harvard), Carter Coberley, James Pope, Aaron Wells (Healthways)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
Overview

- Problem: Model dependence (review)

Gary King (Harvard, IQSS)
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

Lots of insights revealed in the process
- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us compare
Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us compare
- 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 compare
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us compare
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
- \( \leadsto \) Lots of insights revealed in the process
Model Dependence Demonstration

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 Demonstration

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
Model Dependence Demonstration
Replication: Doyle and Sambanis, APSR 2000

- **Data:** 124 Post-World War II civil wars
- **Dependent variable:** peacebuilding success
Model Dependence Demonstration
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)

Gary King (Harvard, IQSS)
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 Demonstration
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?

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
Model Dependence Demonstration
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
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?
<table>
<thead>
<tr>
<th>Variables</th>
<th>Original “Interactive” Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>SE</td>
</tr>
<tr>
<td>Wartype</td>
<td>-1.742</td>
<td>.609</td>
</tr>
<tr>
<td>Logdead</td>
<td>-.445</td>
<td>.126</td>
</tr>
<tr>
<td>Wardur</td>
<td>.006</td>
<td>.006</td>
</tr>
<tr>
<td>Factnum</td>
<td>-1.259</td>
<td>.703</td>
</tr>
<tr>
<td>Factnum2</td>
<td>.062</td>
<td>.065</td>
</tr>
<tr>
<td>Trnsfcap</td>
<td>.004</td>
<td>.002</td>
</tr>
<tr>
<td>Develop</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>Exp</td>
<td>-6.016</td>
<td>3.071</td>
</tr>
<tr>
<td>Decade</td>
<td>-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>

<table>
<thead>
<tr>
<th></th>
<th>Original “Interactive” Model</th>
<th>Modified Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-45.649</td>
<td>-44.902</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>.423</td>
<td>.433</td>
</tr>
</tbody>
</table>
In Sample Fit

Counterfactual Prediction

Probabilities from original model

Probabilities from modified model

Probabilities from original model

Probabilities from modified model
What to do?

Preprocess I: Eliminate extrapolation region

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

Model remaining imbalance

Gary King (Harvard, IQSS)
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)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
What to do?
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
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.

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

Matching reduces model dependence, bias, and variance
What Matching Does

Notation:

Y<br>

i

Dependent variable

T<br>

i

Treatment variable (0/1)

X<br>

i

Pre-treatment covariates

Treatment Effect for treated (T_i = 1) observation:

TE<br>

i

= Y_i (T_i = 1) - Y_i (T_i = 0) = observed - unobserved

Estimate Y_i (0) with Y_j from matched (X_i ≈ 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)
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

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

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

QoI: Sample Average Treatment effect on the Treated:

\( \text{SATT} = \frac{1}{n_T} \sum_{i \in \{ T_i = 1 \}} \text{TE}_i \) or Feasible Average Treatment effect on the Treated: FSATT
What Matching Does

- Notation:
  - $Y_i$: Dependent variable

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

- $Y_i(0)$ 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:
  \[ \text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i \]
  or Feasible Average Treatment effect on the Treated: $\text{FSATT}$

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
What Matching Does

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

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

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

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

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

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

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 \(i\):
  \[
  \text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)
  \]
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*
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**
  \[ \hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j) \]
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**
  \[
  \hat{Y}_i(0) = Y_j(0) \quad \text{or a model} \quad \hat{Y}_i(0) = \hat{g}_0(X_j)
  \]

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

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

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

- **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:**
  \[
  SATT = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} TE_i
  \]

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

Preprocess (Matching)

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

Match each treated unit to the nearest control unit

Control units: not reused; pruned if unused

Prune matches if \( \text{Distance} > \text{caliper} \)

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
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

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

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 $> \text{caliper}$

2. **Estimation** Difference in means or a model
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
Method 3: Coarsened Exact Matching

Preprocess (Matching)

1. Temporarily coarsen X as much as you're willing, e.g., Education (grade school, high school, college, graduate).
2. Apply exact matching to the coarsened X, C(X).
3. Sort observations into strata, each with unique values of C(X).
4. Prune any stratum with 0 treated or 0 control units.
5. Pass on original (uncoarsened) units except those pruned.

Estimation

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

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
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

Gary King (Harvard, IQSS)
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

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

**Bias** (and model dependence) = \( f(\text{imbalance, importance, estimator}) \) ⟷ we measure imbalance instead

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

**Bias-Variance trade off** ⟷ **Imbalance-Variance trade off**

**Measuring Imbalance**

Classic measure: Difference of means (for each variable)

Better measure (difference of multivariate histograms):

\[
L_1(f, g; H) = \frac{1}{2} \sum_{\ell_1, \ldots, \ell_k \in H}(X) |f_{\ell_1, \ldots, \ell_k} - g_{\ell_1, \ldots, \ell_k}|
\]
Bias (& model dependence) = \( f(\text{imbalance}, \text{importance}, \text{estimator}) \)

\( \leadsto \) we measure \text{imbalance} instead
The Bias-Variance Trade Off in Matching

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

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

---

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10

12/27
The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) $= f(\text{imbalance}, \text{importance}, \text{estimator})$
  $\leadsto$ we measure \text{imbalance} instead
- **Variance** $= f(\text{matched sample size}, \text{estimator})$
  $\leadsto$ we measure \text{matched sample size} instead
- **Bias-Variance trade off** $\leadsto$ **Imbalance-}n Trade Off**
The Bias-Variance Trade Off in Matching

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

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

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

- Measuring Imbalance
The Bias-Variance Trade Off in Matching

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

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

- **Bias-Variance trade off** \[\mapsto\] **Imbalance-\( n \) Trade Off

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

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

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

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
Comparing Matching Methods

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

Gary King (Harvard, IQSS)
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

Healthways Data

- Raw Data
- Random Pruning
- PSM
- MDM
- CEM

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
A Space Graph: Real Data

FDA Data

N of Matched Sample ("variance")

L1 ("bias")

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
A Space Graph: Real Data

Lalonde Data Subset

L1 ("bias")

N of Matched Sample ("variance")

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
A Space Graph: Simulated Data — Mahalanobis

MDM: 1 Covariate

- Imbalance:
  - High
  - Med
  - Low

N of matched sample

MDM: 2 Covariates

- Imbalance:
  - High
  - Med
  - Low

N of matched sample

MDM: 3 Covariates

- Imbalance:
  - High
  - Med
  - Low

N of matched sample
A Space Graph: Simulated Data — CEM

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

PSM: 1 Covariate

PSM: 2 Covariates

PSM: 3 Covariates

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
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)
Data where PSM Works Reasonably Well — CEM

Bad CEM: $L_1 = 0.661$

Better CEM: $L_1 = 0.188$

Even Better CEM: $L_1 = 0.095$
CEM Weight: \[ w_i = \frac{m_i^T}{m_i^C} \] (Unnormalized)
CEM Weights and Nonparametric Propensity Score

CEM Weight: \[ w_i = \frac{m_i^T}{m_i^C} \] (Unnormalized)

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: \[ w_i = \frac{m_i^T}{m_i^C} \] (Unnormalized)

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

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

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

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: \[ w_i = \frac{m^T_i}{m^C_i} \] (Unnormalized)

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

⇒ CEM:
- Gives a better pscore than PSM
- Doesn’t match based on crippled information
PSM Approximates Random Matching in Balanced Data

Gary King (Harvard, IQSS)

Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
Destroying CEM with PSM’s Two Step Approach

CEM Matches
CEM-generated PSM Matches

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10
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 is a mistake
- Adjusting experimental data with PSM is a mistake
- Reestimating the propensity score after eliminating noncommon support may be a 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 is a mistake
    - Adjusting experimental data with PSM is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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

Gary King (Harvard, IQSS)
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
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)
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
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 is a mistake
    - Adjusting experimental data with PSM is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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* is a mistake

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 is a mistake
  - Adjusting experimental data with PSM is a mistake

CEM > Mahalanobis > Propensity Score

Reestimating the propensity score after eliminating noncommon support may be a mistake

In four data sets and many simulations:

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* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake

Gary King (Harvard, IQSS)
Talk at Quantitative Issues in Cancer Research Working Seminar, Biostatistics, HSPH, 10/18/10

*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 is a mistake
    - Adjusting experimental data with PSM is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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 is a mistake
    - Adjusting experimental data with PSM is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a 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 and Stata), tutorials, etc.

http://GKing.Harvard.edu/cem