Matching for Causal Inference Without Balance Checking

Gary King
Institute for Quantitative Social Science
Harvard University

joint work with
Stefano M. Iacus (Univ. of Milan) and Giuseppe Porro (Univ. of Trieste)

(talk at the University of Rochester, 1/19/09)

Preview Slide: Coarsened Exact Matching (CEM)

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) Temporarily coarsen X as much as you're willing

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) Temporarily coarsen X as much as you're willing
   - e.g., Education (grade school, high school, college, graduate)

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on the coarsened X, $C(X)$

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on the coarsened X, $C(X)$
   - Sort observations into strata, each with unique values of $C(X)$

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:

   (A) 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

   (B) Perform exact matching on 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

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

   (Or apply other matching methods within CEM strata
   & they inherert CEM's properties)

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

   (Or apply other matching methods within CEM strata
   & they inherert CEM's properties)

$\rightsquigarrow$ A version of CEM: Last studied 40 years ago by Cochran

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

   (Or apply other matching methods within CEM strata
   & they inherert CEM's properties)

⇝ A version of CEM: Last studied 40 years ago by Cochran

⇝ First used many decades before that

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

   (Or apply other matching methods within CEM strata
   & they inherert CEM's properties)

$\rightsquigarrow$ A version of CEM: Last studied 40 years ago by Cochran

$\rightsquigarrow$ First used many decades before that

$\rightsquigarrow$ We prove: many new properties, uses, & extensions,

Preview Slide: Coarsened Exact Matching (CEM)

A simple (and ancient) method of causal inference, with surprisingly powerful properties

1. Preprocess (X, T) with CEM:
   (A) 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

   (B) Perform exact matching on 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

   (C) Pass on original (uncoarsened) units except those pruned
2. Analyze as without matching (adding weights for stratum-size)

   (Or apply other matching methods within CEM strata
   & they inherert CEM's properties)

⇝ A version of CEM: Last studied 40 years ago by Cochran

⇝ First used many decades before that

⇝ We prove: many new properties, uses, & extensions,
and show how it resolves many problems in the literature

Characteristics of Observational Data

Characteristics of Observational Data

- Lots of data

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment:

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random,

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator,

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment:
  not random, not controlled by investigator, not known

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment:
  not random, not controlled by investigator, not known
- Bias-Variance Tradeoff

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias
- Removing heterogeneous data will often reduce variance too

Characteristics of Observational Data

- Lots of data
- Data is of uncertain origin. Treatment assignment: not random, not controlled by investigator, not known
- Bias-Variance Tradeoff
- The idea of matching: sacrifice some data to avoid bias
- Removing heterogeneous data will often reduce variance too
- (Medical experiments are the reverse: small-n with random treatment assignment; don't match unless something goes wrong)

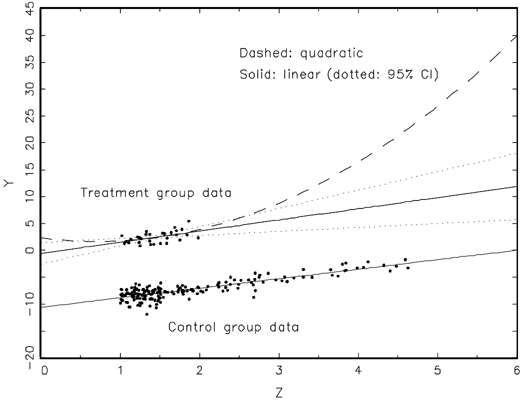
Model Dependence

Model Dependence

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

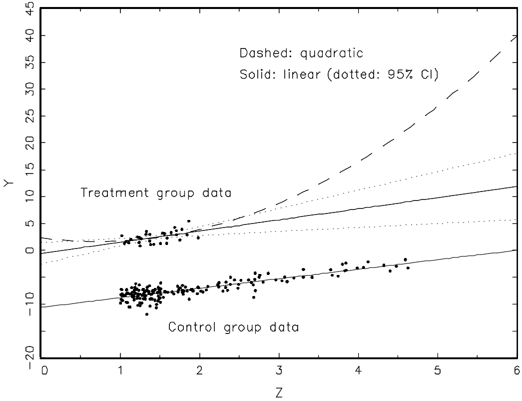
Model Dependence

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



Model Dependence

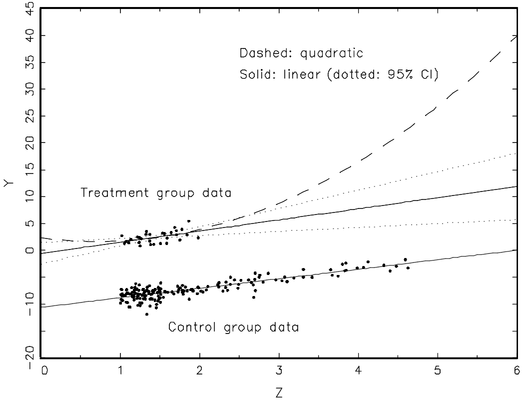
(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

Model Dependence

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

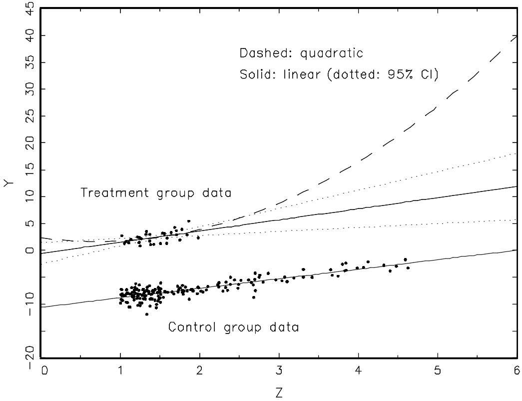


What to do?

- Preprocess I: Eliminate extrapolation region (a separate step)

Model Dependence

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

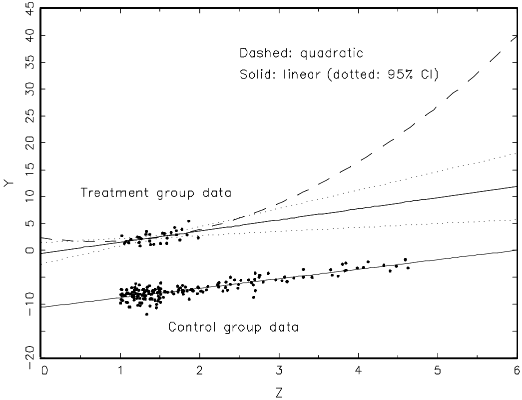


What to do?

- Preprocess I: Eliminate extrapolation region (a separate step)
- Preprocess II: Match (prune bad matches) within interpolation region

Model Dependence

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



What to do?

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

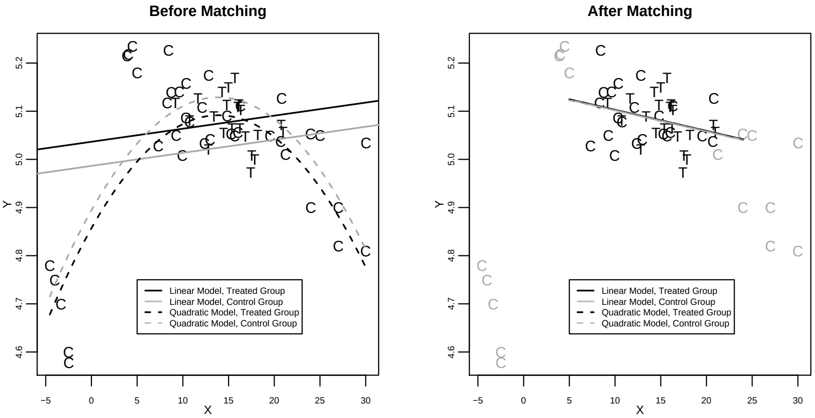
Matching within the Interpolation Region

Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

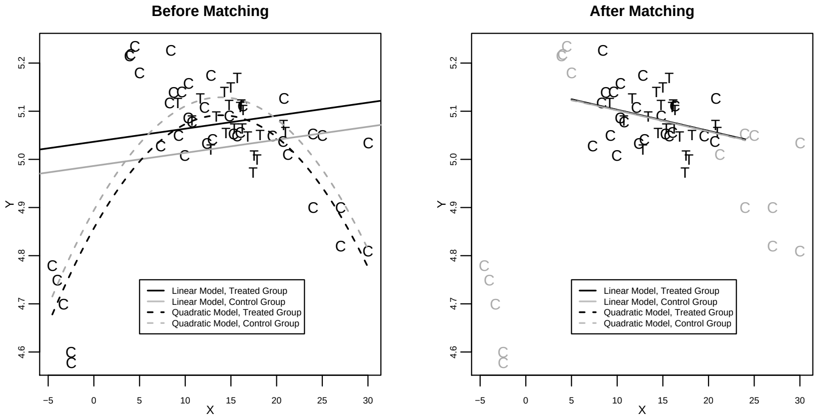
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching reduces model dependence, bias, and variance

The Goals, with some more precision

The Goals, with some more precision

- Notation:

The Goals, with some more precision

- Notation:

  Y_i Dependent variable

The Goals, with some more precision

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)

The Goals, with some more precision

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1)
  X_i pre-treatment covariates

The Goals, with some more precision

- 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:

The Goals, with some more precision

- 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:

$$\text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

The Goals, with some more precision

- 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:

$$\begin{aligned}\text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved}\end{aligned}$$

The Goals, with some more precision

- 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:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

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

The Goals, with some more precision

- 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:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- 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)

The Goals, with some more precision

- 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:

$$\begin{aligned}\text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved}\end{aligned}$$

- 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)
- Sample Average Treatment effect on the Treated:

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

Problems With Existing Matching Methods

Problems With Existing Matching Methods

- Don't eliminate extrapolation region

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance):

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES);

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y;

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X;

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance;

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
    - Least important (variance): matched n chosen ex ante
    - Most important (bias): imbalance reduction checked ex post

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\leadsto$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice:

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
    - Least important (variance): matched n chosen ex ante
    - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
    - Best practice: choose n-match

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
    - Least important (variance): matched n chosen ex ante
    - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
    - Best practice: choose n-match-check, tweak-match

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\leadsto$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice:

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose n,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
    - Least important (variance): matched n chosen ex ante
    - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
    - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
    - Actual practice: choose n, match,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose n, match, publish,

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose n, match, publish, STOP.

Problems With Existing Matching Methods

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched n chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose n-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose n, match, publish, STOP. (Is balance even improved?)

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched *n* checked ex post

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched *n* checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \gamma(\epsilon)$	**vars to adjust**
$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon)$	remaining vars

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \gamma(\epsilon) \qquad \textbf{vars to adjust}$$

$$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon) \qquad \text{remaining vars}$$

Treated and control X variables to adjust

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \gamma(\epsilon) \qquad \textbf{vars to adjust}$$

$$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon) \qquad \text{remaining vars}$$

Remaining treated and control X variables

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$$D(\mathbf{X}^{\epsilon}_{T}, \mathbf{X}^{\epsilon}_{C}) \leq \gamma(\epsilon) \qquad \textbf{vars to adjust}$$

$$D(X^{\epsilon}_{T}, X^{\epsilon}_{C}) \leq \gamma(\epsilon) \qquad \text{remaining vars}$$

"Imbalance" given chosen distance metric

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \gamma(\boldsymbol{\epsilon}) \qquad \textbf{vars to adjust}$$

$$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon) \qquad \text{remaining vars}$$

Bounds (maximum imbalance)

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \gamma(\boldsymbol{\epsilon})$	**vars to adjust**
$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon)$	remaining vars

One tuning parameter ϵ_j, one for each X_j

A New Class of Methods: Monotonic Imbalance Bounding

- No restrictions on data types
- Designed for observational data (reversing EPBR):
  - Most important (bias): degree of balance chosen ex ante
  - Least important (variance): matched n checked ex post
- Balance is measured in sample (like blocked designs), not merely in expectation (like complete randomization)
- Covers all forms of imbalance: means, interactions, nonlinearities, moments, multivariate histograms, etc.
- One adjustable tuning parameter per variable
- Convenient monotonicity property: Reducing maximum imbalance on one X: no effect on others

MIB Formally (simplifying for this talk):

$D(\mathbf{X}_T^{\epsilon}, \mathbf{X}_C^{\epsilon}) \leq \boldsymbol{\gamma}(\boldsymbol{\epsilon})$	**vars to adjust**
$D(X_T^{\epsilon}, X_C^{\epsilon}) \leq \gamma(\epsilon)$	remaining vars

If $\boldsymbol{\epsilon}$ is reduced, $\boldsymbol{\gamma}(\boldsymbol{\epsilon})$ decreases & $\gamma(\epsilon)$ is unchanged

What's Coarsening?

What's Coarsening?

- Coarsening is intrinsic to measurement

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
  - Education: grade school, middle school, high school, college, graduate

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
  - Education: grade school, middle school, high school, college, graduate
  - Income: poverty level threshold, or larger bins for higher income

What's Coarsening?

- Coarsening is intrinsic to measurement
  - We think of measurement as clarity between categories
  - But measurement also involves homogeneity within categories
  - Examples: male/female, rich/middle/poor, black/white, war/nonwar.
  - Better measurement devices (e.g., telescopes) produce more detail
- Data analysts routinely coarsen, thinking grouping error is less risky than measurement error. E.g.:
  - 7 point Party ID $\rightsquigarrow$ Democrat/Independent/Republican
  - Likert Issue questions $\rightsquigarrow$ agree/{neutral,no opinion}/disagree
  - multiparty voting $\rightsquigarrow$ winner/losers
  - Religion, Occupation, SEC industries, ICD codes, etc.
- Temporary Coarsening for CEM; e.g.:
  - Education: grade school, middle school, high school, college, graduate
  - Income: poverty level threshold, or larger bins for higher income
  - Age: infant, child, adolescent, young adult, middle age, elderly

CEM as an MIB Method

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for:

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles,

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set . . .

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set . . .
  - too large?

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set ...
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set ...
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances
  - too small?

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set . . .
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances
  - too small? $\rightsquigarrow$ n may be too small

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set . . .
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances
  - too small? $\rightsquigarrow$ n may be too small
  - as large as you're comfortable with, but n is still too small?

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set ...
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances
  - too small? $\rightsquigarrow$ n may be too small
  - as large as you're comfortable with, but n is still too small?
    $\rightsquigarrow$ No magic method of matching can save you;

CEM as an MIB Method

- Define: ϵ as largest (coarsened) bin size ($\epsilon = 0$ is exact matching)
- We Prove: setting ϵ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.
  $\Longrightarrow$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)
- By default, both treated and control units are pruned: CEM estimates a quantity that can be estimated without model dependence
- What if ϵ is set ...
  - too large? $\rightsquigarrow$ You're left modeling remaining imbalances
  - too small? $\rightsquigarrow$ n may be too small
  - as large as you're comfortable with, but n is still too small?
    $\rightsquigarrow$ No magic method of matching can save you;
    $\rightsquigarrow$ You're stuck modeling or collecting better data

Other CEM properties we prove

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
    - The principle: data space = analysis space
    - Estimators that violate it are nonrobust and counterintuitive

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit):

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student;

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

	CEM	pscore	Mahalanobis	Genetic
% Common Units	96.5	70.2	80.9	80.0

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

	CEM	pscore	Mahalanobis	Genetic
% Common Units	96.5	70.2	80.9	80.0

- Fast and memory-efficient even for large n; can be fully automated

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: ϵ_j is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

	CEM	pscore	Mahalanobis	Genetic
% Common Units	96.5	70.2	80.9	80.0

- Fast and memory-efficient even for large n; can be fully automated
- Simple to teach: coarsen, then exact match

Imbalance Measures

Imbalance Measures

Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

Imbalance Measures

Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

Multivariate Imbalance: difference in histograms (bins fixed ex ante)

$$\mathcal{L}_1(f, g) = \sum_{\ell_1 \cdots \ell_k} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

Imbalance Measures

Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

Multivariate Imbalance: difference in histograms (bins fixed ex ante)

$$\mathcal{L}_1(f, g) = \sum_{\ell_1 \cdots \ell_k} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

Local Imbalance by Variable (given strata fixed by matching method)

$$I_2^{(j)} = \frac{1}{S} \sum_{s=1}^{S} \left| \bar{X}_{m_T^s}^{(j)} - \bar{X}_{m_C^s}^{(j)} \right|, \quad j = 1, \ldots, k$$

CEM in Practice: EPBR-Compliant Data

CEM in Practice: EPBR-Compliant Data

Monte Carlo:

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$.

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$
initial						1.24

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$
initial						1.24
PSC	2.38	1.25	.74	1.25	.74	1.18

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$
initial						1.24
PSC	2.38	1.25	.74	1.25	.74	1.18
MAH	.56	.36	.29	.36	.29	1.13

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$
initial						1.24
PSC	2.38	1.25	.74	1.25	.74	1.18
MAH	.56	.36	.29	.36	.29	1.13
CEM	.42	.26	.17	.22	.19	.78

CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means (I_1):

	X_1	X_2	X_3	X_4	X_5	Seconds
initial	1.00	1.00	1.00	1.00	1.00	
MAH	.20	.20	.20	.20	.20	.28
PSC	.11	.06	.03	.06	.03	.16
CEM	.04	.02	.06	.06	.04	.08

Local (I_2) and multivariate $\mathcal{L}_1$ imbalance:

	X_1	X_2	X_3	X_4	X_5	$\mathcal{L}_1$
initial						1.24
PSC	2.38	1.25	.74	1.25	.74	1.18
MAH	.56	.36	.29	.36	.29	1.13
CEM	.42	.26	.17	.22	.19	.78

$\rightsquigarrow$ CEM dominates EPBR-methods in EPBR Data

CEM in Practice: Non-EPBR Data

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	-423.7	1566.5	1622.6	.00	1.28

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	−423.7	1566.5	1622.6	.00	1.28
MAH	784.8	737.9	1077.2	.03	1.08

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	−423.7	1566.5	1622.6	.00	1.28
MAH	784.8	737.9	1077.2	.03	1.08
PSC	260.5	1025.8	1058.4	.02	1.23

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	−423.7	1566.5	1622.6	.00	1.28
MAH	784.8	737.9	1077.2	.03	1.08
PSC	260.5	1025.8	1058.4	.02	1.23
GEN	78.3	499.5	505.6	27.38	1.12

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	−423.7	1566.5	1622.6	.00	1.28
MAH	784.8	737.9	1077.2	.03	1.08
PSC	260.5	1025.8	1058.4	.02	1.23
GEN	78.3	499.5	505.6	27.38	1.12
CEM	.8	111.4	111.4	.03	.76

CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

	BIAS	SD	RMSE	Seconds	$\mathcal{L}_1$
initial	−423.7	1566.5	1622.6	.00	1.28
MAH	784.8	737.9	1077.2	.03	1.08
PSC	260.5	1025.8	1058.4	.02	1.23
GEN	78.3	499.5	505.6	27.38	1.12
CEM	.8	111.4	111.4	.03	.76

⇝ CEM works well in non-EPBR data too

CEM Extensions I

CEM Extensions I

- CEM and Multiple Imputation for Missing Data

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
    1. put missing observation in stratum where plurality of imputations fall
    2. pass on uncoarsened imputations to analysis stage
    3. Use the usual MI combining rules to analyze

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata
- Automating user choices

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata
- Automating user choices Histogram bin size calculations,

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound,

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound, Progressive Coarsening

CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory treatments: No modification necessary; keep all strata with ≥ 1 unit having each value of T
- Blocking in Randomized Experiments: no modification needed; randomly assign T within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound, Progressive Coarsening
- Detecting Extreme Counterfactuals

CEM Extensions II: Improving Existing Matching Methods

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)
3. Mahalanobis distance: can apply within CEM strata

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)
3. Mahalanobis distance: can apply within CEM strata
4. Genetic Matching: can constrain results to CEM strata

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)
3. Mahalanobis distance: can apply within CEM strata
4. Genetic Matching: can constrain results to CEM strata
5. Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)
3. Mahalanobis distance: can apply within CEM strata
4. Genetic Matching: can constrain results to CEM strata
5. Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness
6. Nonparametric Adjustments: can apply within CEM strata

CEM Extensions II: Improving Existing Matching Methods

1. Most commonly used methods:
   - cannot be used to eliminate extrapolation region
   - don't possess most other CEM properties
   - but inherent CEM properties if applied within CEM strata
2. Propensity Score matching:
   - requires correct specification or balance can drop (the usual specification tests are irrelevant; must check balance)
   - CEM strata can bound bias in pscore matching
   - may be good for applications with many covariates we know little about (so we're willing to take balance on any subset)
3. Mahalanobis distance: can apply within CEM strata
4. Genetic Matching: can constrain results to CEM strata
5. Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness
6. Nonparametric Adjustments: can apply within CEM strata
7. $\rightsquigarrow$ & whatever else you all come up with

For papers, software (for R and Stata), tutorials, etc.

http://GKing.Harvard.edu/cem