Simplifying Matching Methods for Causal Inference[1]

Gary King[2]

Institute for Quantitative Social Science
Harvard University

(Talk at the U.S. Food and Drug Administration [CBER/OBE], 8/25/2015)

[1]Based on joint work with Rich Nielsen, Chris Lucas, Stefano Iacus, and Giuseppe Porro

[2]GaryKing.org

Part 1 (of 3)

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher Discretion $\rightsquigarrow$ Bias

Model Dependence Example

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

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

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,. . .
- Counterfactual question: Switch UN intervention for each war

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000
(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,...
- Counterfactual question: Switch UN intervention for each war
- Data analysis: Logit model

Model Dependence Example

Replication of Doyle and Sambanis, APSR 2000

(From: King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,. . .
- Counterfactual question: Switch UN intervention for each war
- Data analysis: Logit model
- The question: How *model dependent* are the results?

Two Logit Models, Apparently Similar Results

Two Logit Models, Apparently Similar Results

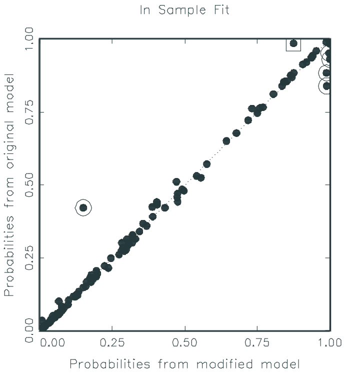
Effect of Multilateral UN Intervention on Peacebuilding Success

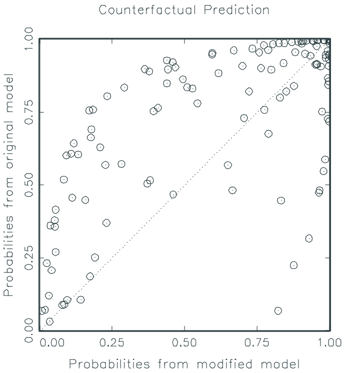
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

	Original "Interactive" Model			Modified Model		
Variables	Coeff	SE	P-val	Coeff	SE	P-val
Wartype	−1.742	.609	.004	−1.666	.606	.006
Logdead	−.445	.126	.000	−.437	.125	.000
Wardur	.006	.006	.258	.006	.006	.342
Factnum	−1.259	.703	.073	−1.045	.899	.245
Factnum2	.062	.065	.346	.032	.104	.756
Trnsfcap	.004	.002	.010	.004	.002	.017
Develop	.001	.000	.065	.001	.000	.068
Exp	−6.016	3.071	.050	−6.215	3.065	.043
Decade	−.299	.169	.077	−0.284	.169	.093
Treaty	2.124	.821	.010	2.126	.802	.008
UNOP4	3.135	1.091	.004	.262	1.392	.851
Wardur*UNOP4	—	—	—	.037	.011	.001
Constant	8.609	2.157	0.000	7.978	2.350	.000
N		122			122	
Log-likelihood		-45.649			-44.902	
Pseudo R^2		.423			.433	

Model Dependence: Same Fit, Different Predictions





Coarsened Exact Matching

Coarsened Exact Matching (CEM)

Coarsened Exact Matching (CEM)

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

Coarsened Exact Matching (CEM)

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

1. Preprocess (X, T) with CEM:

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

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)

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 a histogram

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 a histogram

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

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 a histogram

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

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 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

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 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

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 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)

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 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)

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 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 45 years ago by Cochran

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 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 45 years ago by Cochran
$\rightsquigarrow$ First used many decades before that

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 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 45 years ago by Cochran

$\rightsquigarrow$ First used many decades before that

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

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 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 45 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

Problems with Other Matching Methods

Problems with Other Matching Methods

- Don't eliminate the extrapolation region

Problems with Other Matching Methods

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

Problems with Other Matching Methods

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

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data:

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES);

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y;

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X;

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance;

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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:

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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:

Problems with Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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 Other Matching Methods

- Don't eliminate the extrapolation region
- Don't work with multiply imputed data
- Violate the congruence principle
- Matching methods from the largest class (EPBR, e.g., PSM, MDM) don't apply to real data: require normal data (or DMPES); all X's must have same effect on Y; Y must be a linear function of X; aim 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?)

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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution

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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 measurements (e.g., telescopes) $\rightsquigarrow$ better resolution
- 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 a Monotonic Imbalance Bounding Method

CEM as a Monotonic Imbalance Bounding Method

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

CEM as a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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.
  $\implies$ Setting ϵ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched n is determined ex post)

CEM as a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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 a Monotonic Imbalance Bounding 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, researcher discretion, bias, estimation error

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds: model dependence, researcher discretion, bias, estimation error
- Meets the congruence principle

Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds: model dependence, researcher discretion, bias, 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, researcher discretion, bias, 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, researcher discretion, bias, 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, researcher discretion, bias, 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, researcher discretion, bias, 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, researcher discretion, bias, 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, researcher discretion, bias, 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	PSM	MDM	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, researcher discretion, bias, 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	PSM	MDM	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, researcher discretion, bias, 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	PSM	MDM	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

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 MDM & PSM 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 MDM & PSM 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 MDM & PSM to match with replacement;use automated CEM

Difference in means

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 MDM & PSM to match with replacement;use automated CEM

Difference in means

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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	Seconds
initial	1.00	1.00	1.00	1.00	1.00	.50	

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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	Seconds
initial	1.00	1.00	1.00	1.00	1.00	.50	
MDM	.45	.45	.45	.45	.45	.34	.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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	Seconds
initial	1.00	1.00	1.00	1.00	1.00	.50	
MDM	.45	.45	.45	.45	.45	.34	.28
PSM	.32	.32	.32	.32	.32	.31	.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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	Seconds
initial	1.00	1.00	1.00	1.00	1.00	.50	
MDM	.45	.45	.45	.45	.45	.34	.28
PSM	.32	.32	.32	.32	.32	.31	.16
CEM	.04	.04	.08	.06	.07	.21	.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 MDM & PSM to match with replacement;use automated CEM

	Difference in means						
	X_1	X_2	X_3	X_4	X_5	L_1	Seconds
initial	1.00	1.00	1.00	1.00	1.00	.50	
MDM	.45	.45	.45	.45	.45	.34	.28
PSM	.32	.32	.32	.32	.32	.31	.16
CEM	.04	.04	.08	.06	.07	.21	.08

$\rightsquigarrow$ 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.

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	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	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	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	L_1
initial	−423.7	1566.5	1622.6	.00	1.28
MDM	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	L_1
initial	−423.7	1566.5	1622.6	.00	1.28
MDM	784.8	737.9	1077.2	.03	1.08
PSM	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	L_1
initial	−423.7	1566.5	1622.6	.00	1.28
MDM	784.8	737.9	1077.2	.03	1.08
PSM	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	L_1
initial	-423.7	1566.5	1622.6	.00	1.28
MDM	784.8	737.9	1077.2	.03	1.08
PSM	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	L_1
initial	-423.7	1566.5	1622.6	.00	1.28
MDM	784.8	737.9	1077.2	.03	1.08
PSM	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

$\rightsquigarrow$ CEM works well in non-EPBR data too

CEM Extensions I

CEM Extensions I

- Many binary variables:

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
    1. Multiply Impute (missing) or Overimpute (measurement error)

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. Use the usual MI combining rules to analyze

CEM Extensions I

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Many binary variables: $\rightsquigarrow$ coarsen sums of related vars
- Missing Data and/or Measurement Error
  1. Multiply Impute (missing) or Overimpute (measurement error)
  2. Put missing observation in stratum where plurality of imputations fall
  3. Pass on uncoarsened imputations to analysis stage
  4. 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

- Most commonly used methods:

CEM Extensions II: Improving Existing Matching Methods

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

CEM Extensions II: Improving Existing Matching Methods

- 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

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

CEM Extensions II: Improving Existing Matching Methods

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

CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM

CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn't be used in practice

CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn't be used in practice
- MDM: can apply within CEM strata

CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn't be used in practice
- MDM: can apply within CEM strata
- Genetic Matching: can constrain results to CEM strata

CEM Extensions II: Improving Existing Matching Methods

- Most commonly used methods:
    - cannot be used to eliminate extrapolation region
    - don't possess most other CEM properties
    - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
    - CEM strata can bound bias in PSM
    - Probably shouldn't be used in practice
- MDM: can apply within CEM strata
- Genetic Matching: can constrain results to CEM strata
- Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness

CEM Extensions II: Improving Existing Matching Methods

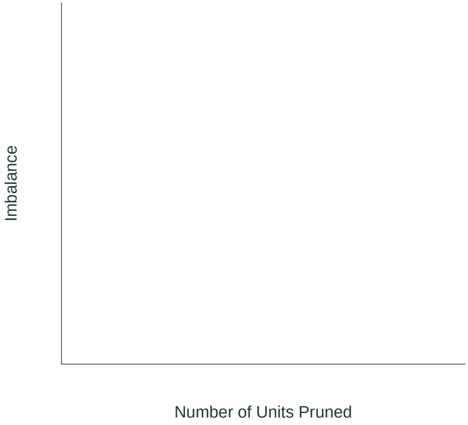
- Most commonly used methods:
  - cannot be used to eliminate extrapolation region
  - don't possess most other CEM properties
  - inherent CEM properties if applied within CEM strata
- Propensity Score matching:
  - CEM strata can bound bias in PSM
  - Probably shouldn't be used in practice
- MDM: can apply within CEM strata
- Genetic Matching: can constrain results to CEM strata
- Synthetic Matching, or Robins' weights: CEM can identify region to apply weights, increasing efficiency/robustness
- Nonparametric Adjustments: can apply within CEM strata

Part 3 (of 3)

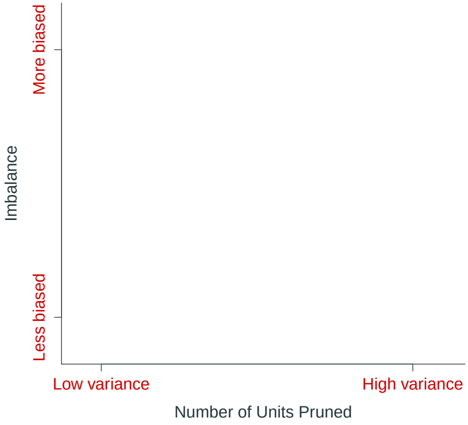
The Matching Frontier

The Matching Frontier

The Matching Frontier



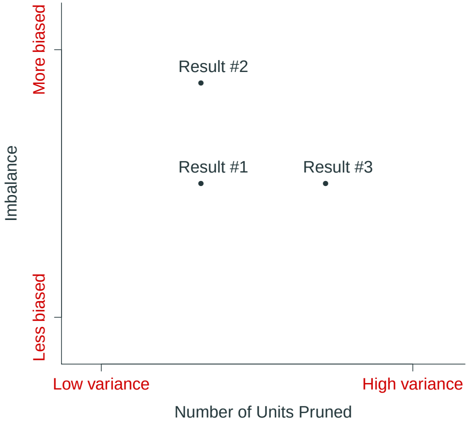
The Matching Frontier



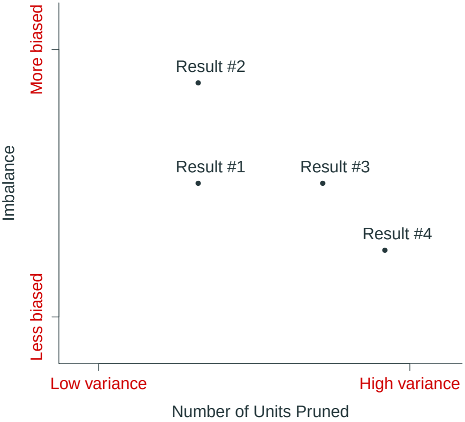
The Matching Frontier

The Matching Frontier

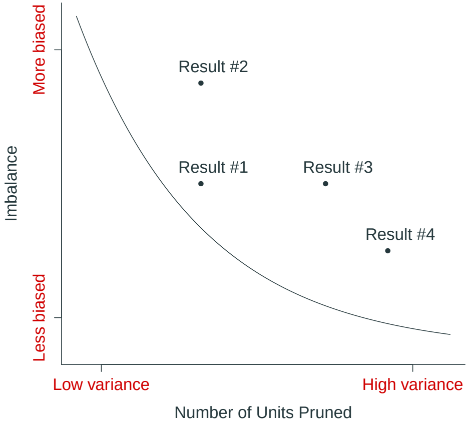
The Matching Frontier



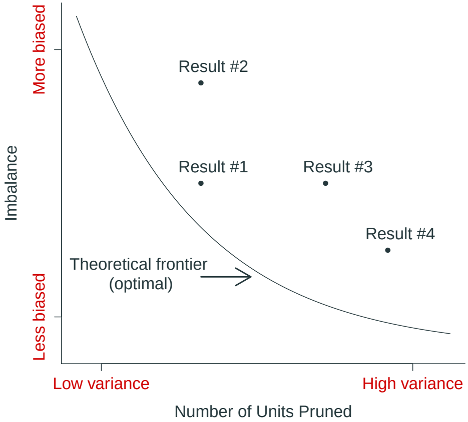
The Matching Frontier



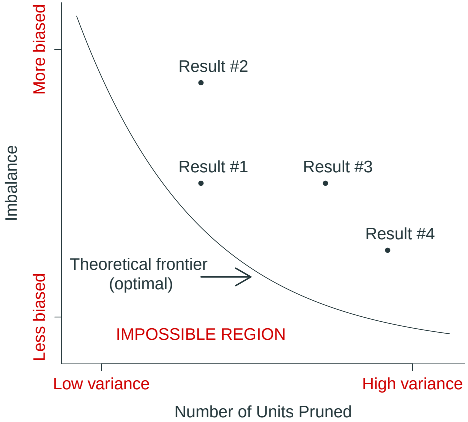
The Matching Frontier



The Matching Frontier



The Matching Frontier



How hard is the frontier to calculate?

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)
  - $\rightsquigarrow$ It's **easy** to calculate!

The Matching Frontier Algorithm

The Matching Frontier Algorithm

- To use, make 2 choices:

The Matching Frontier Algorithm

- To use, make 2 choices:
    1. Quantity of interest: SATT (prune Cs only) or FSATT

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method
  - All solutions are optimal

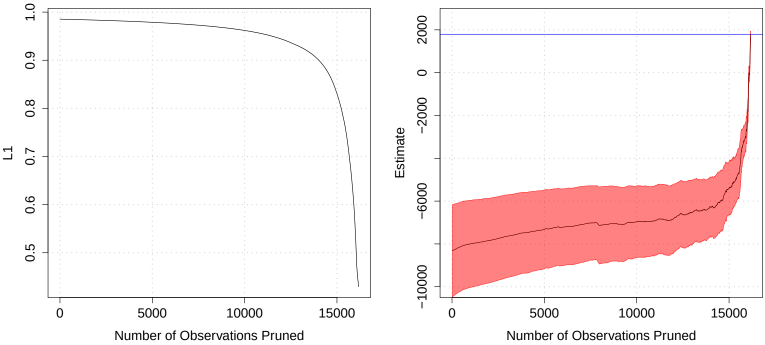
The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method
  - All solutions are optimal
  - No iteration or diagnostics required

The Matching Frontier Algorithm

- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Algorithm finds the whole frontier
  - Simple to use (free easy software available)
  - No need to choose or use a matching method
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible; you see everything optimal

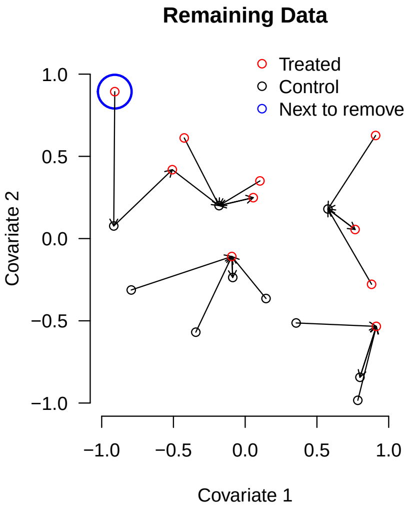
Job Training Data: Frontier and Causal Estimates

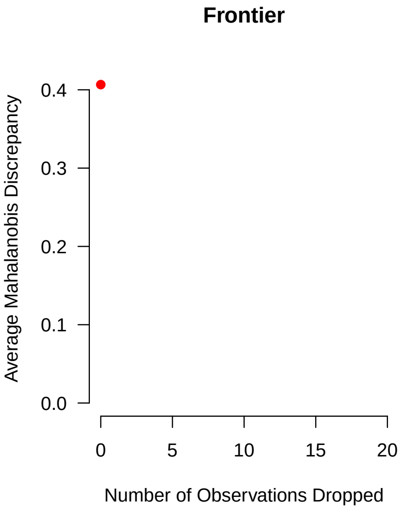


- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after pruning most Cs
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed

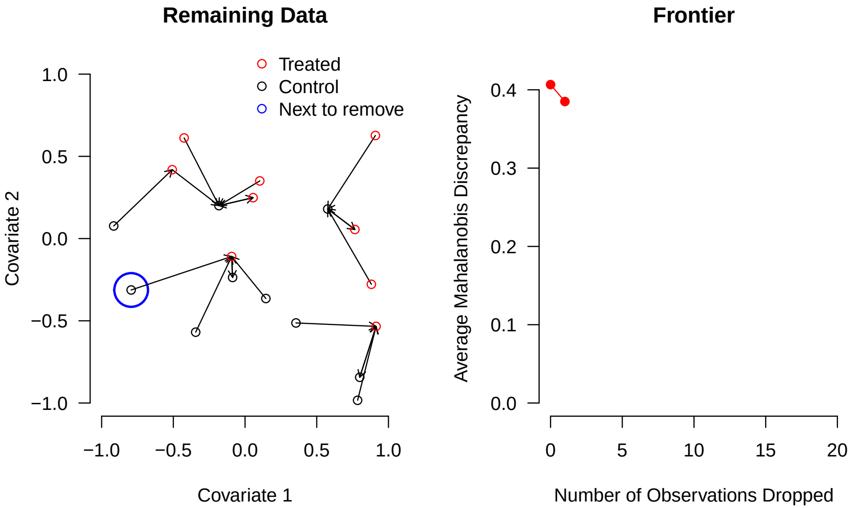
Constructing the FSATT Mahalanobis Frontier

Constructing the FSATT Mahalanobis Frontier

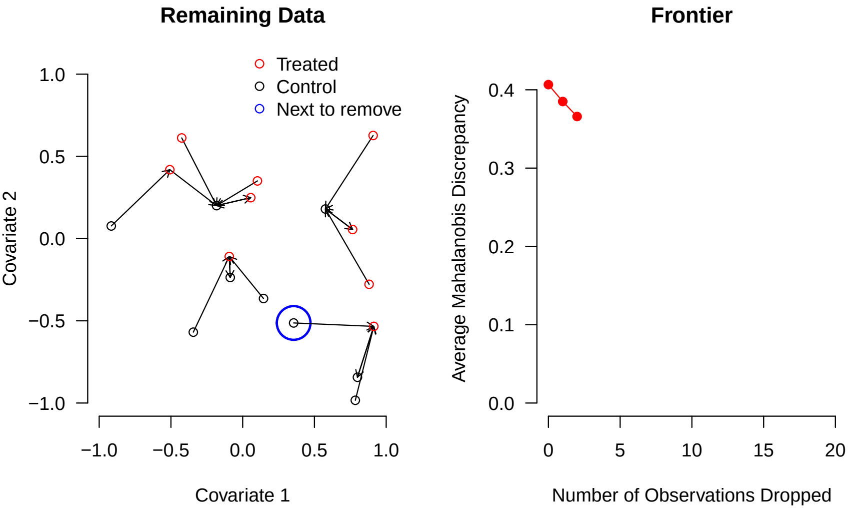




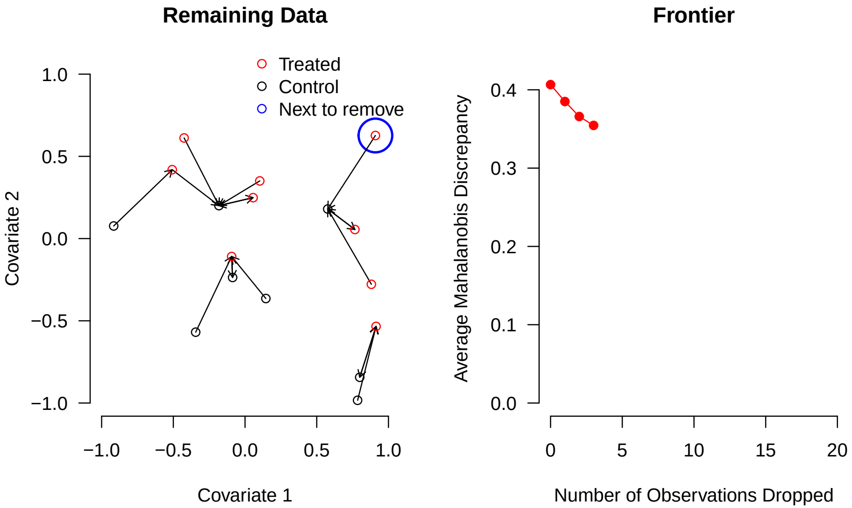
Constructing the FSATT Mahalanobis Frontier



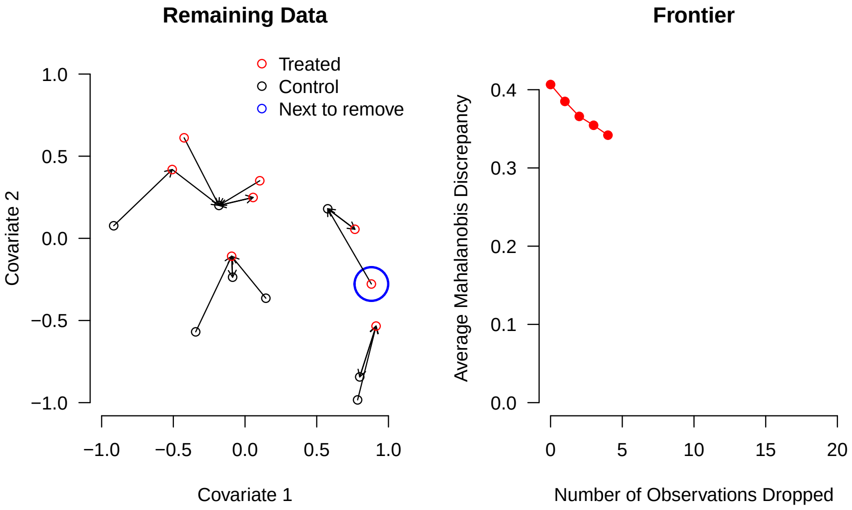
Constructing the FSATT Mahalanobis Frontier



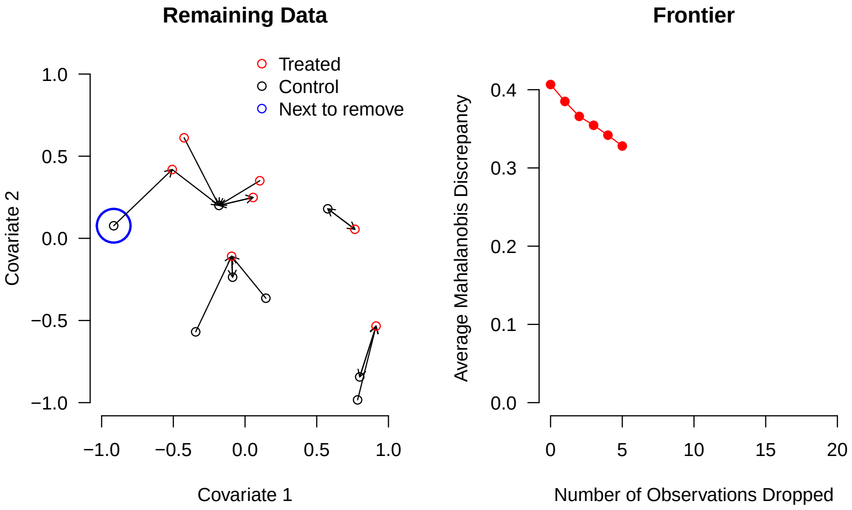
Constructing the FSATT Mahalanobis Frontier



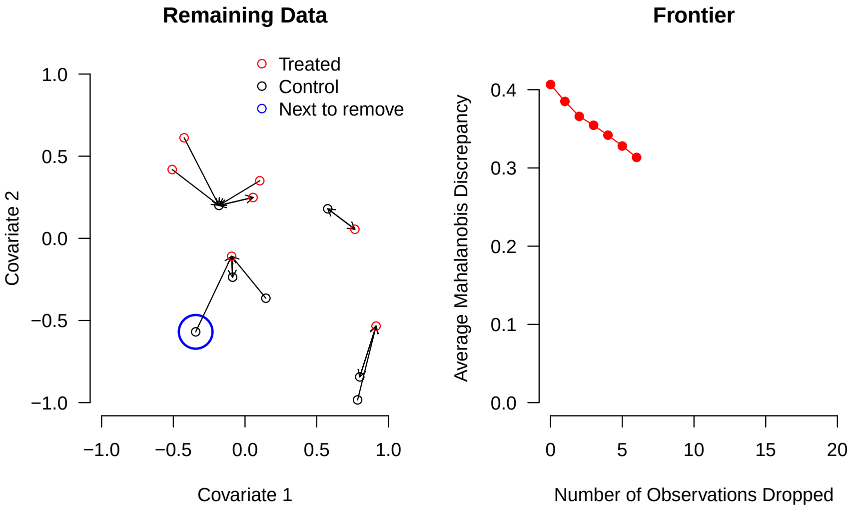
Constructing the FSATT Mahalanobis Frontier



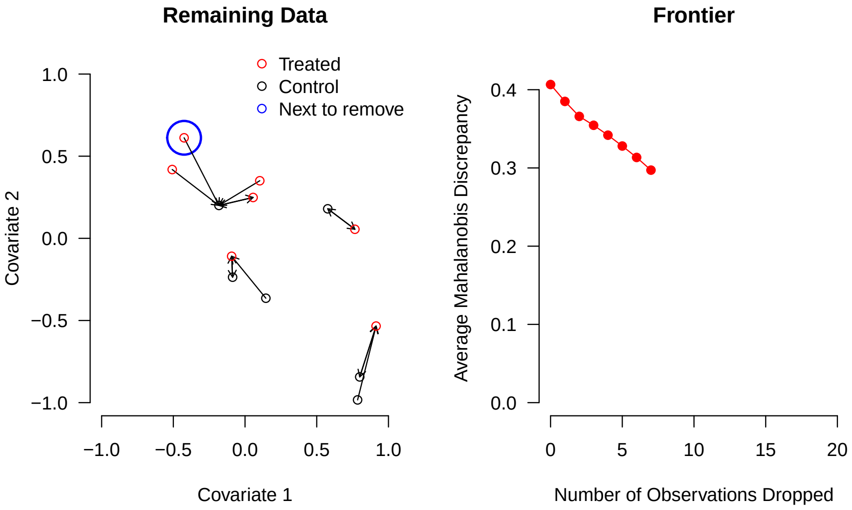
Constructing the FSATT Mahalanobis Frontier



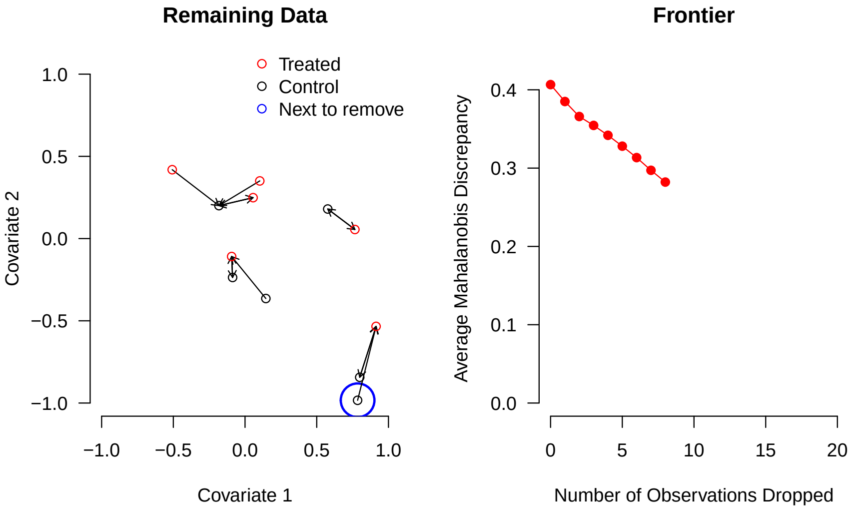
Constructing the FSATT Mahalanobis Frontier



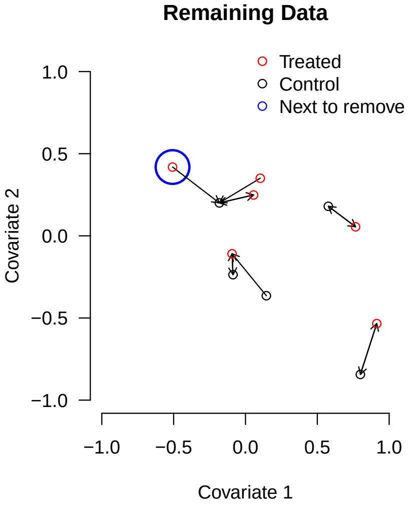
Constructing the FSATT Mahalanobis Frontier

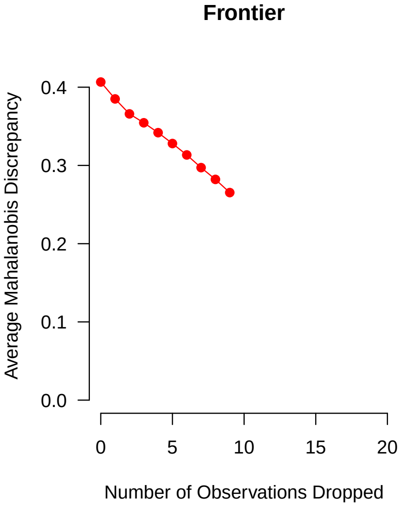


Constructing the FSATT Mahalanobis Frontier

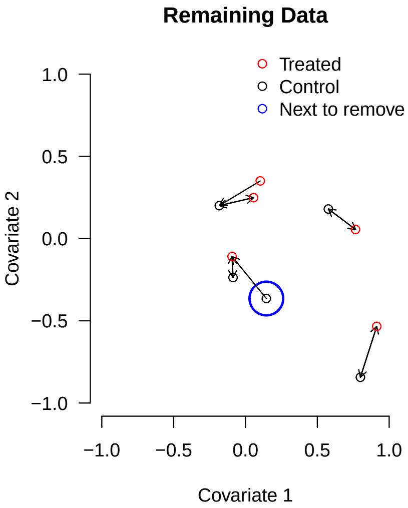


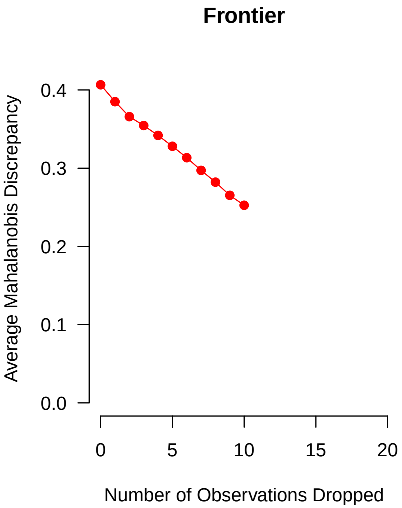
Constructing the FSATT Mahalanobis Frontier



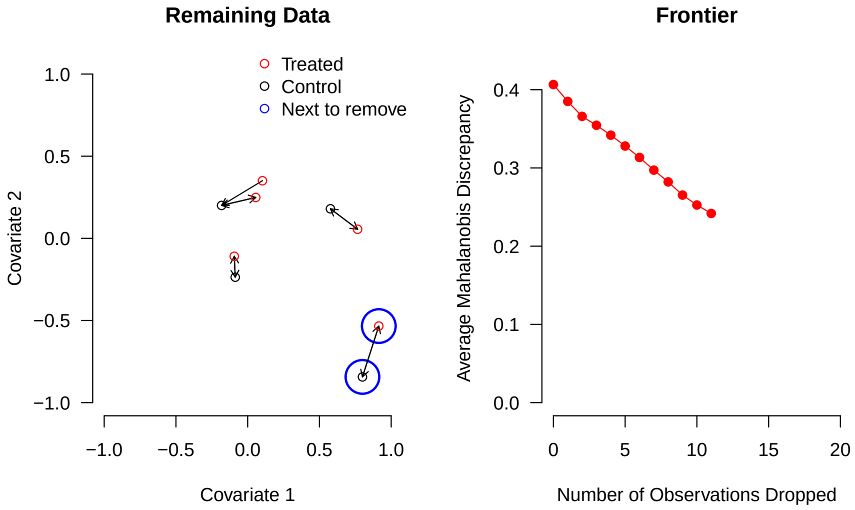


Constructing the FSATT Mahalanobis Frontier

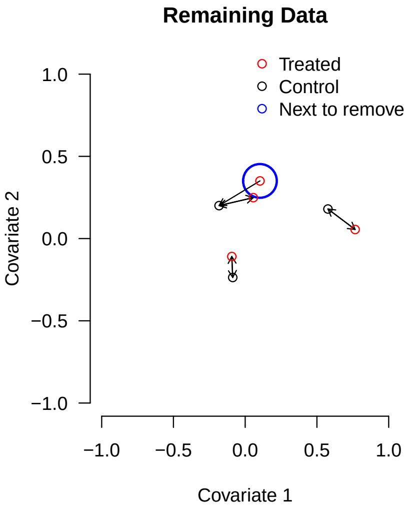


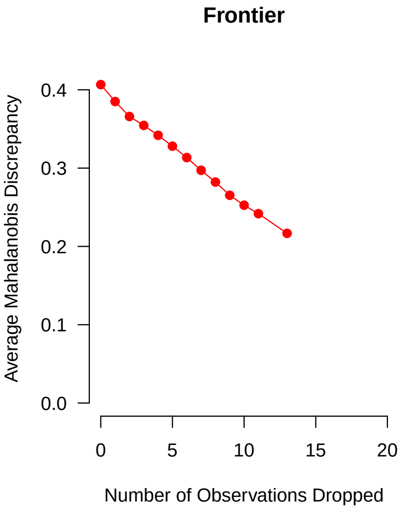


Constructing the FSATT Mahalanobis Frontier

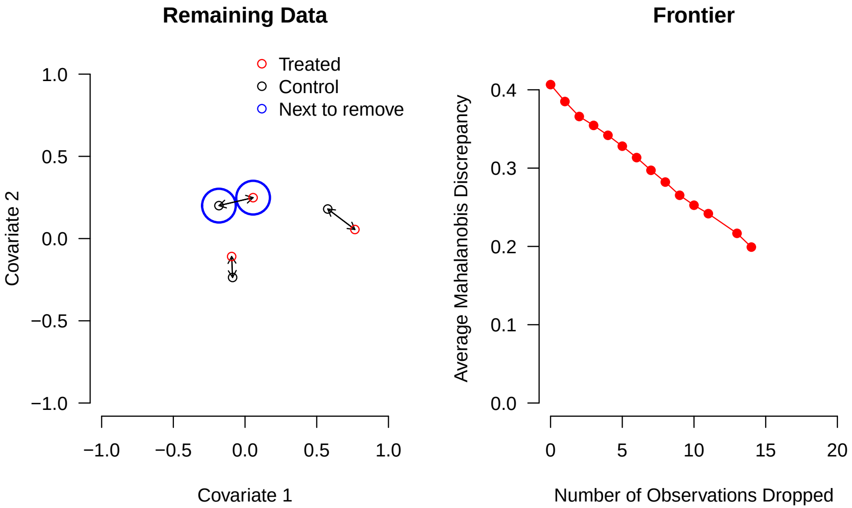


Constructing the FSATT Mahalanobis Frontier

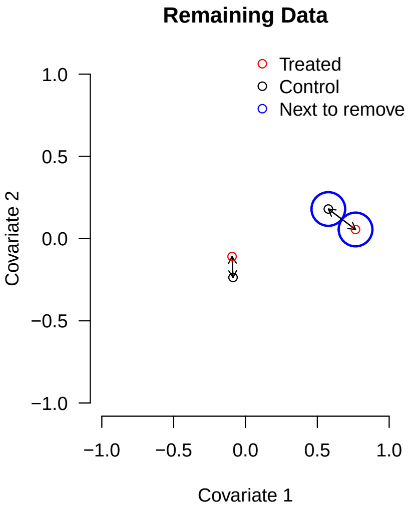


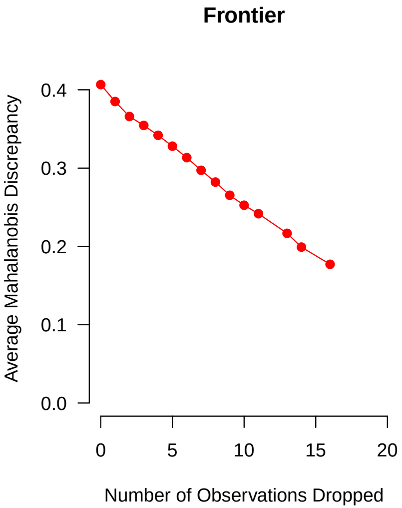


Constructing the FSATT Mahalanobis Frontier

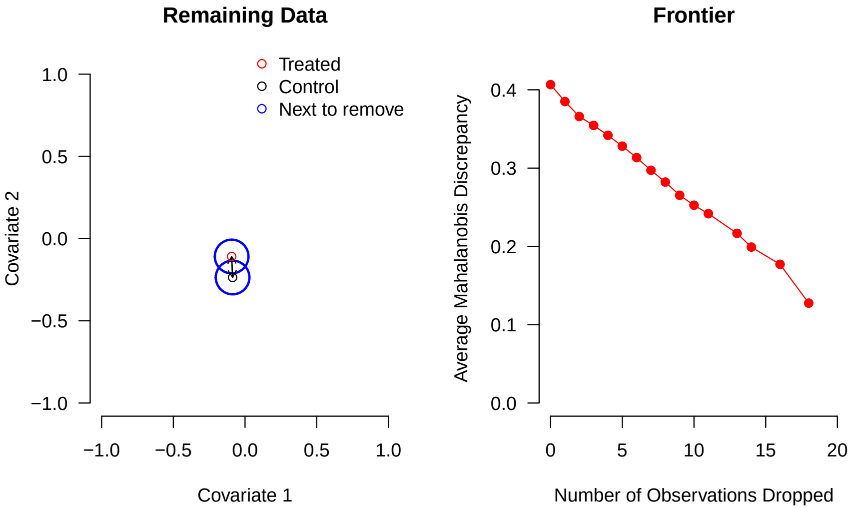


Constructing the FSATT Mahalanobis Frontier

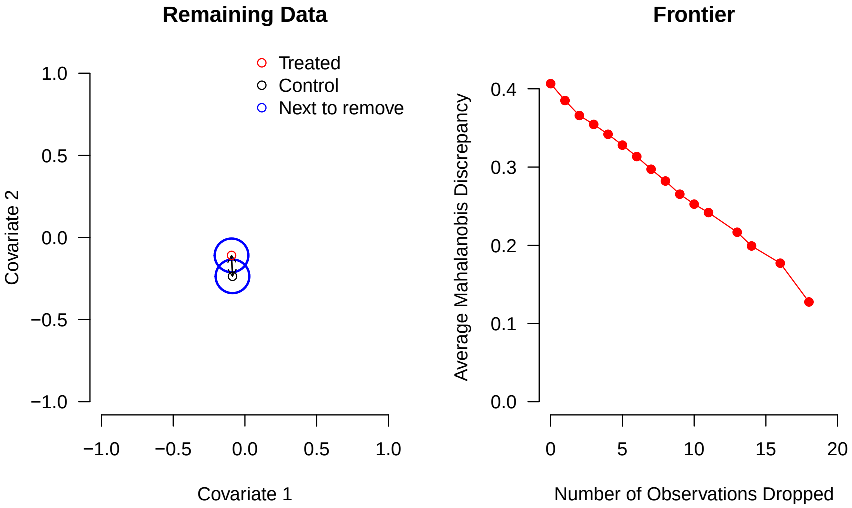




Constructing the FSATT Mahalanobis Frontier

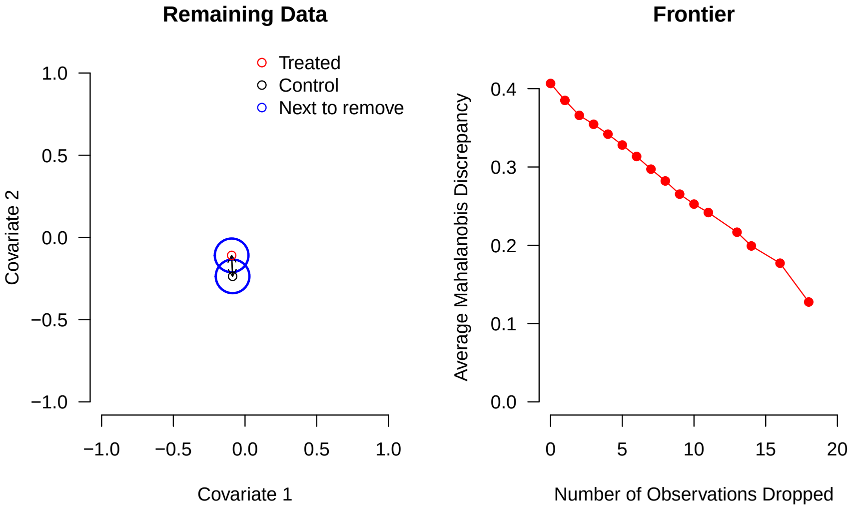


Constructing the FSATT Mahalanobis Frontier



- Warning: figure omits details and the proof!

Constructing the FSATT Mahalanobis Frontier



- Warning: figure omits details and the proof!
- Very fast; works with any continuous imbalance metric

Constructing the L1/L2 SATT Frontier

Constructing the L1/L2 SATT Frontier

Short version:

Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins

Constructing the L1/L2 SATT Frontier

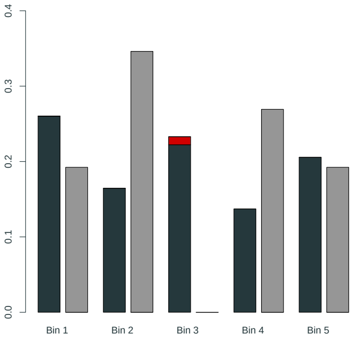
Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

Constructing the L1/L2 SATT Frontier

Short version:

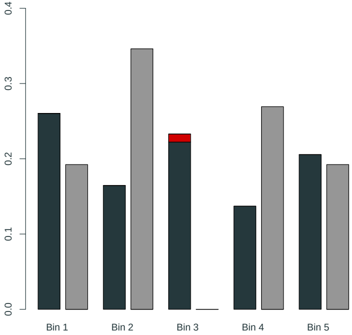
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

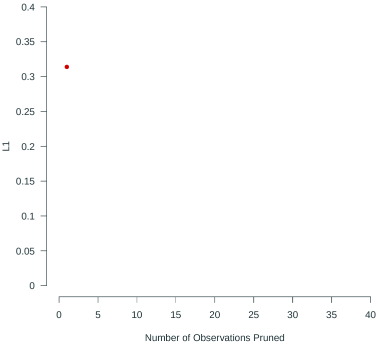


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

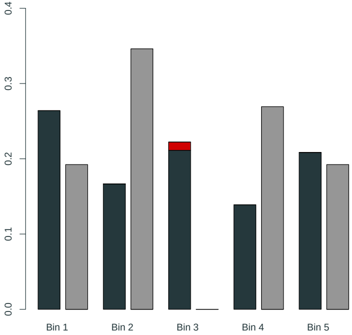


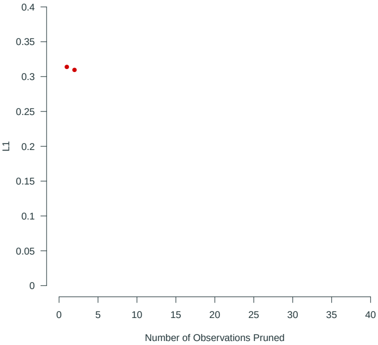


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

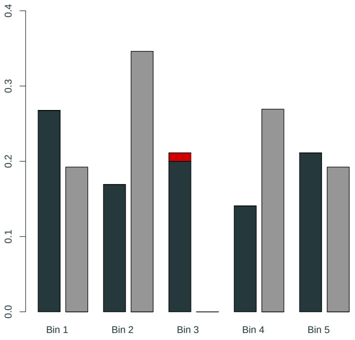


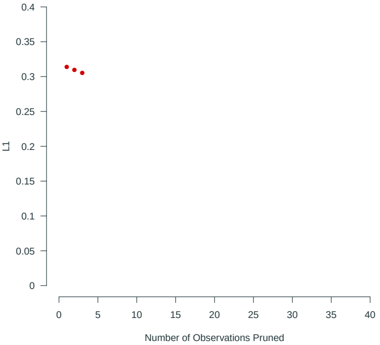


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

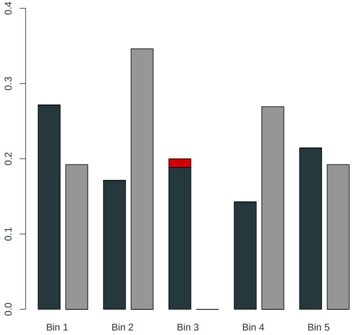


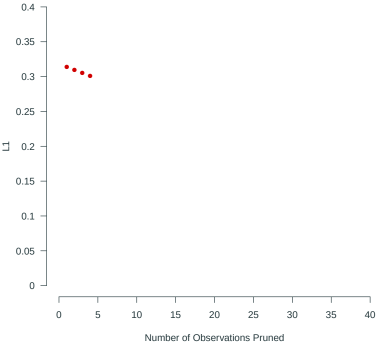


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

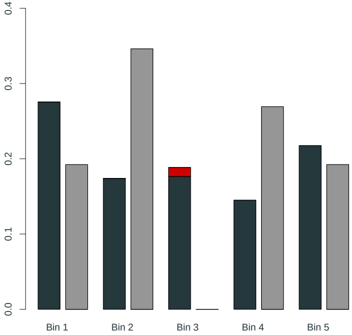


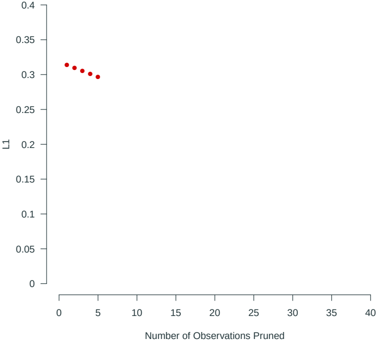


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

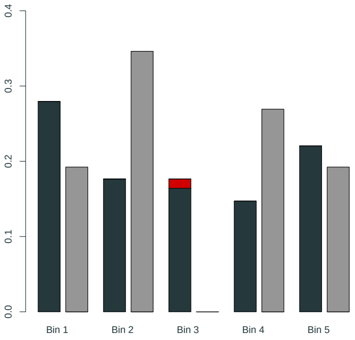


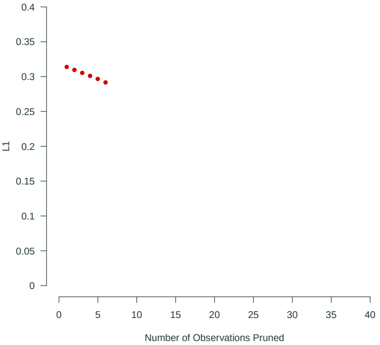


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

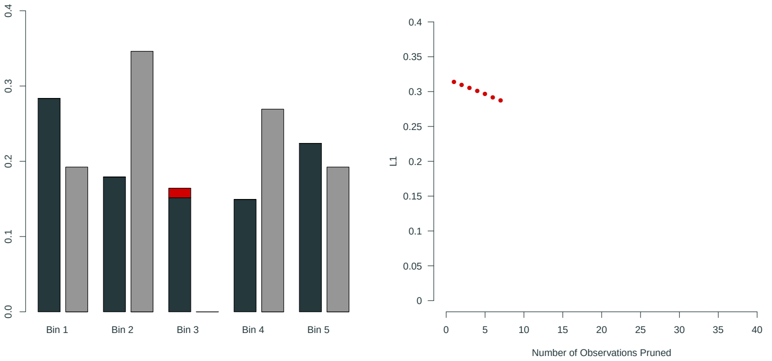




Constructing the L1/L2 SATT Frontier

Short version:

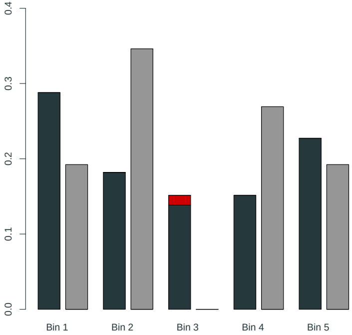
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

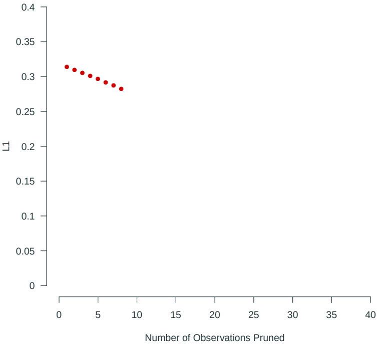


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

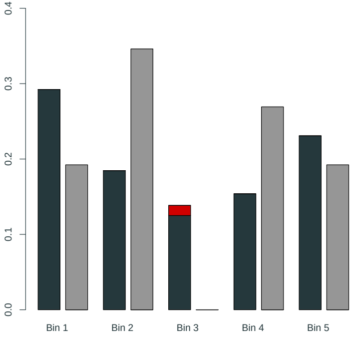


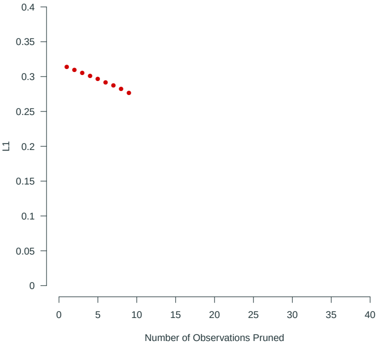


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

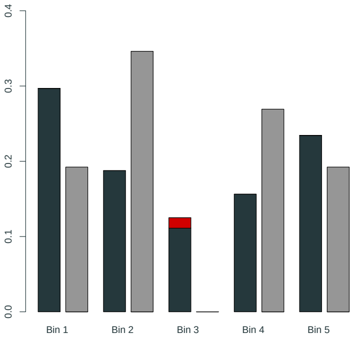


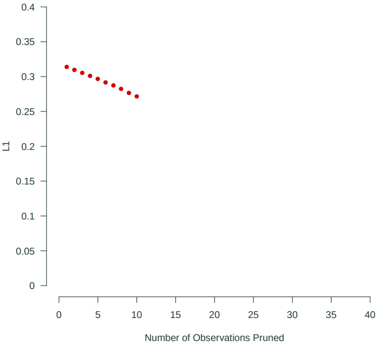


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

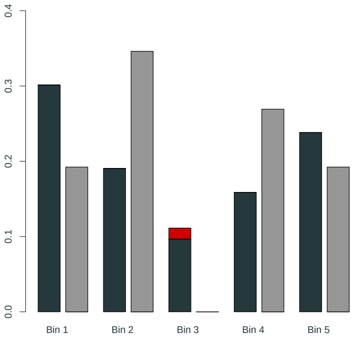


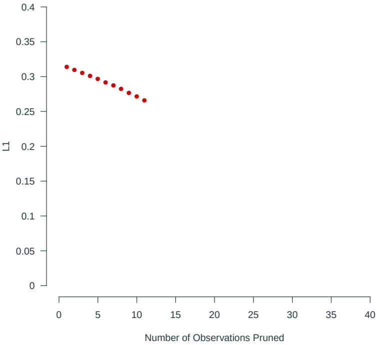


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

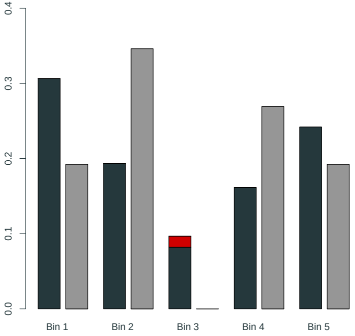


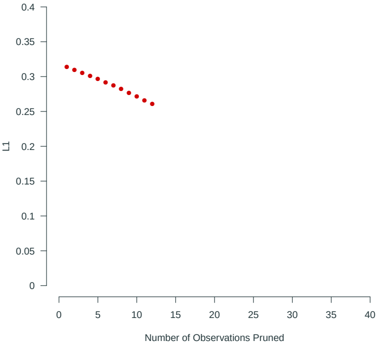


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

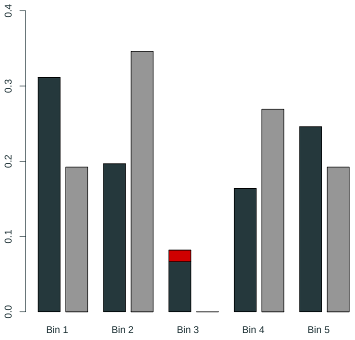


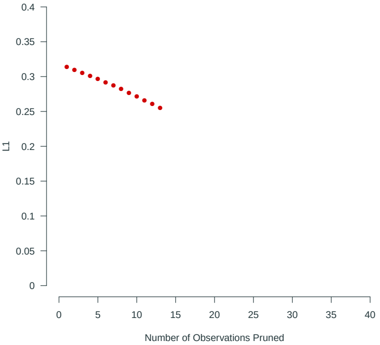


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

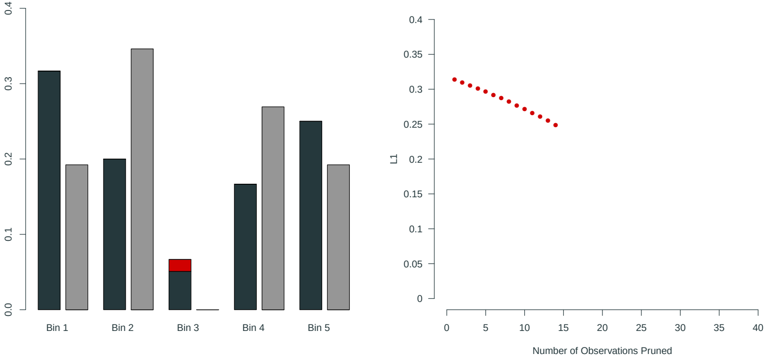




Constructing the L1/L2 SATT Frontier

Short version:

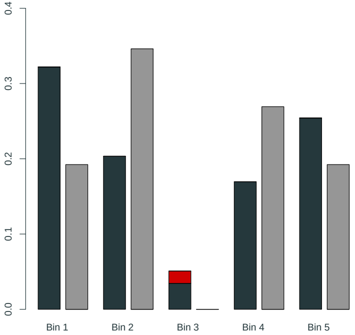
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

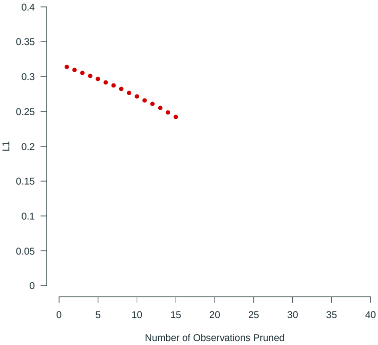


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

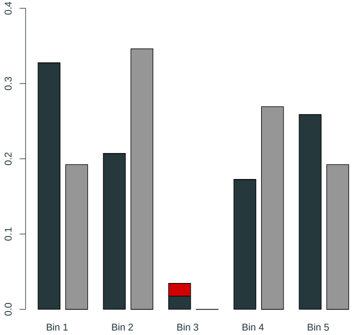


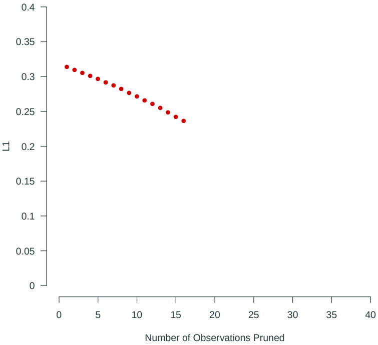


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

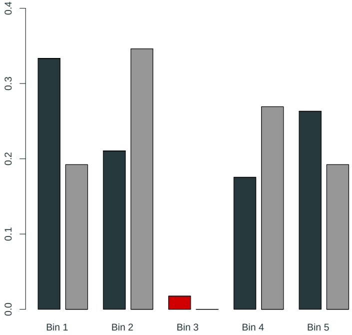


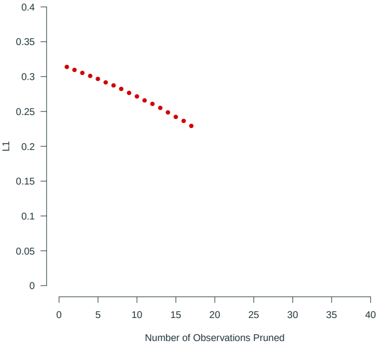


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

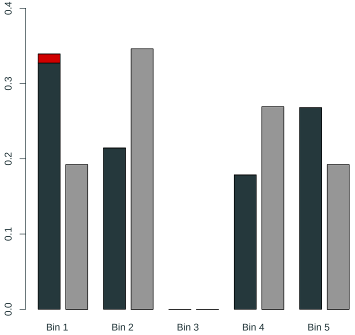


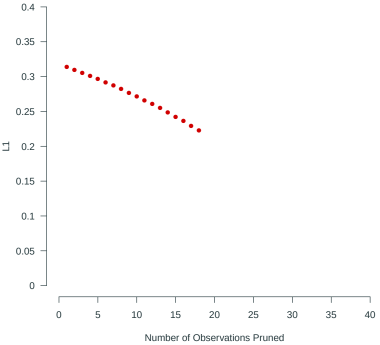


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

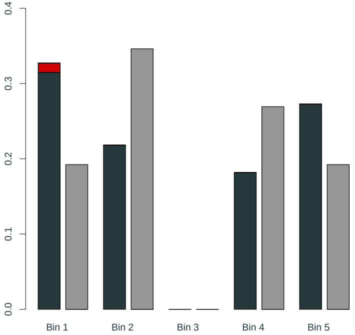


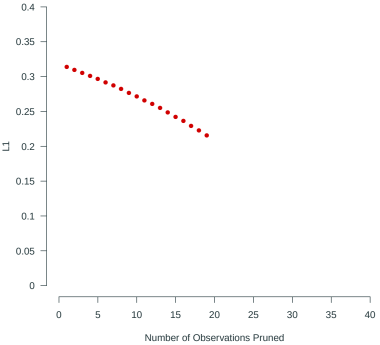


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

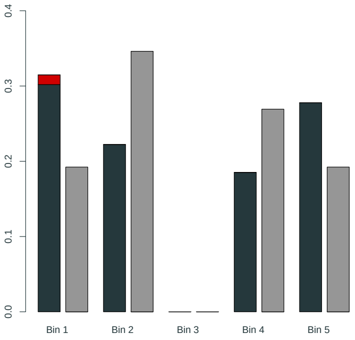


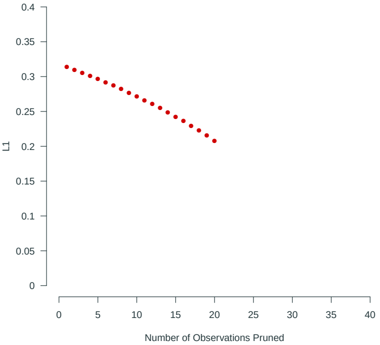


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

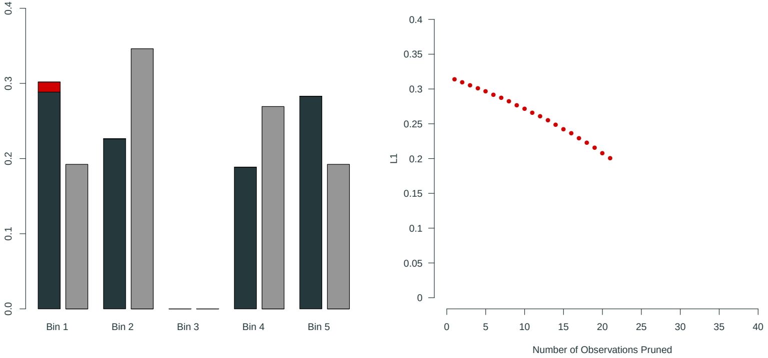




Constructing the L1/L2 SATT Frontier

Short version:

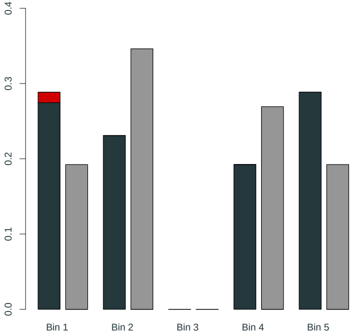
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

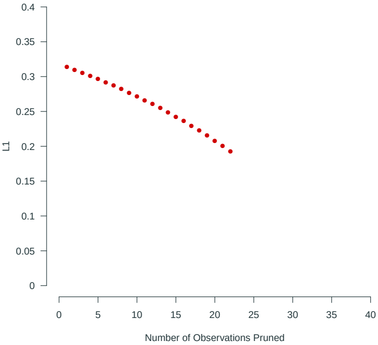


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

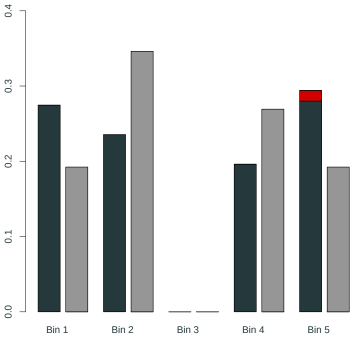


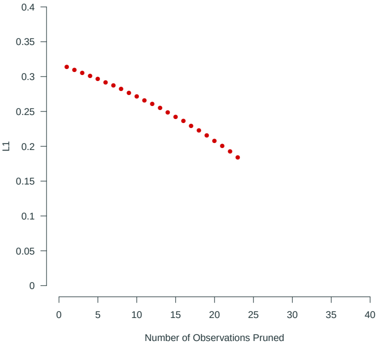


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

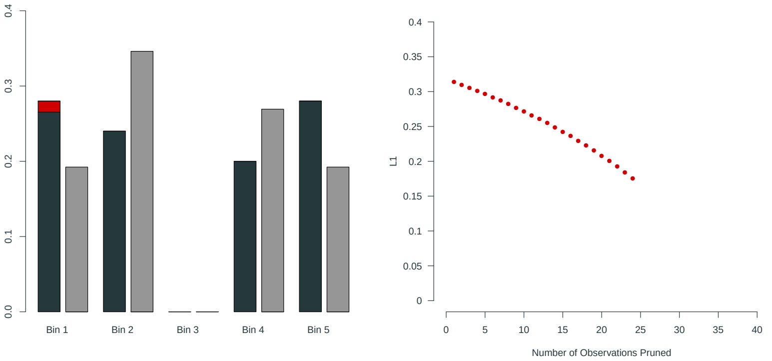




Constructing the L1/L2 SATT Frontier

Short version:

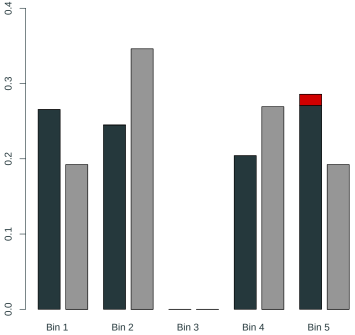
1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

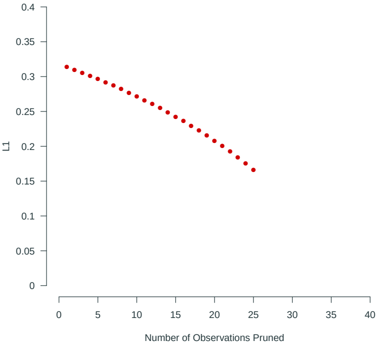


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

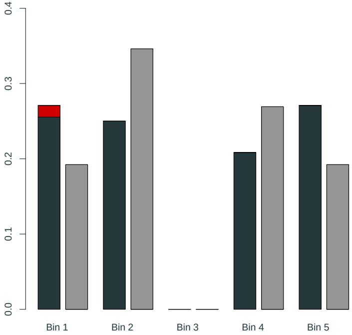


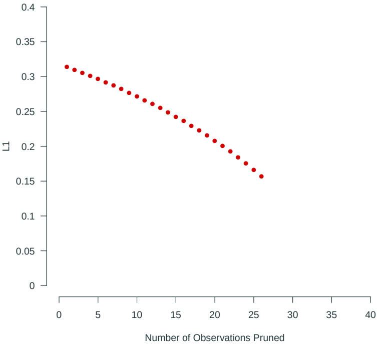


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

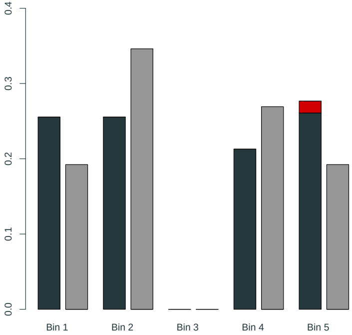


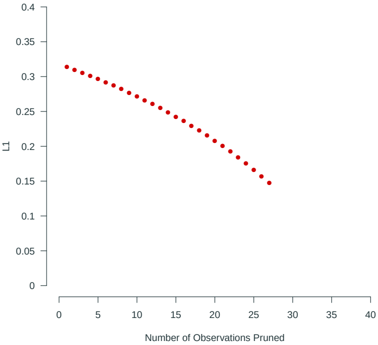


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

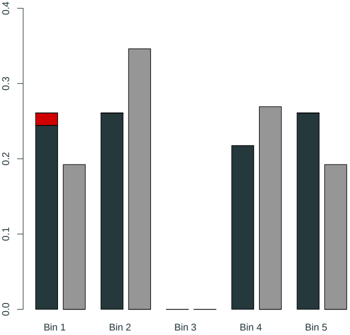


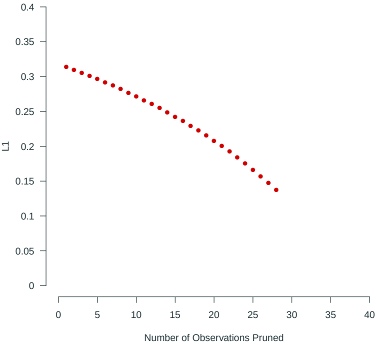


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

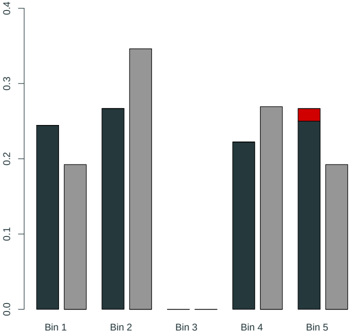


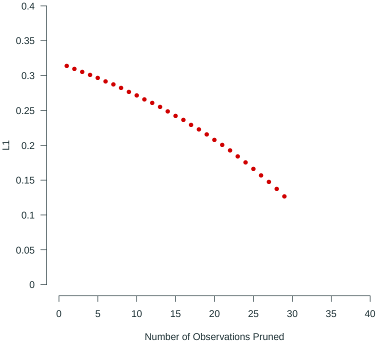


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

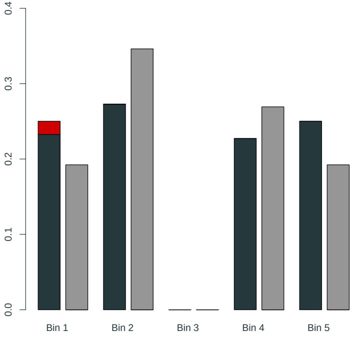


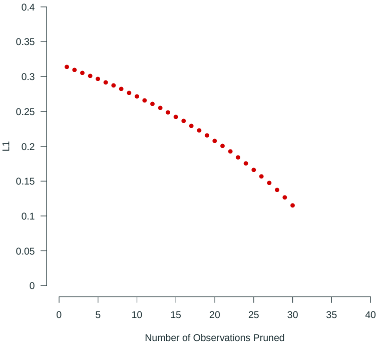


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

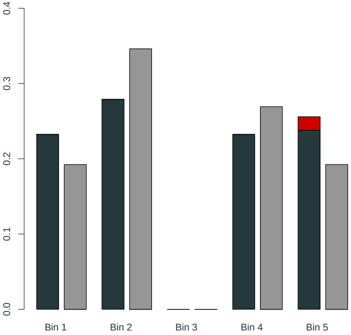


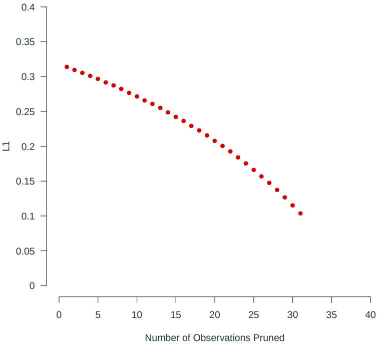


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

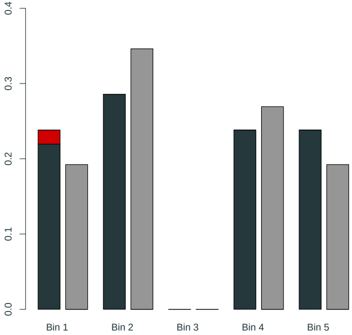


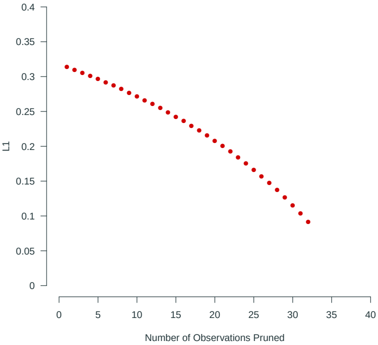


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

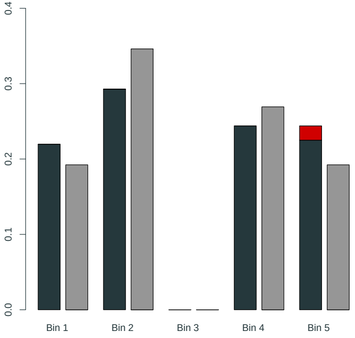


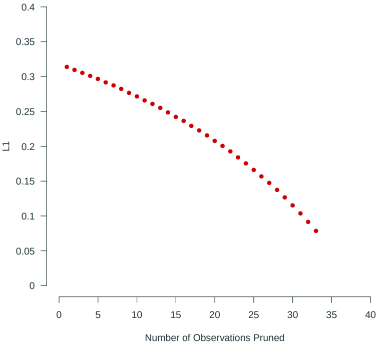


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

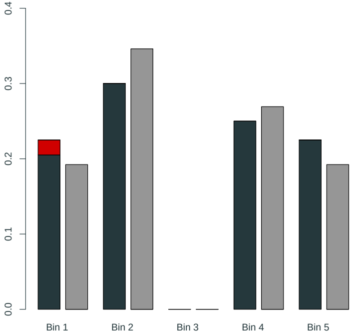


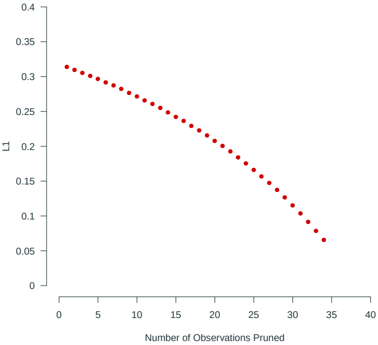


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

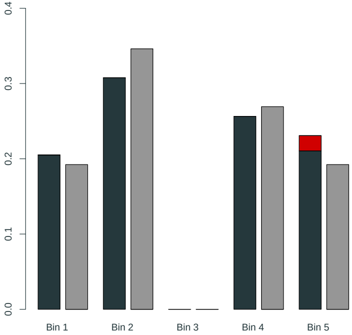


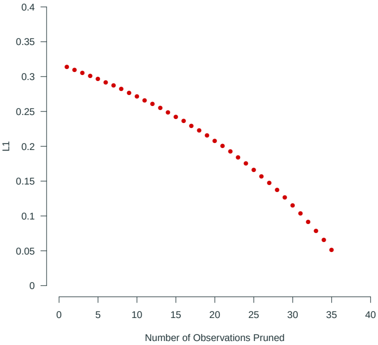


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

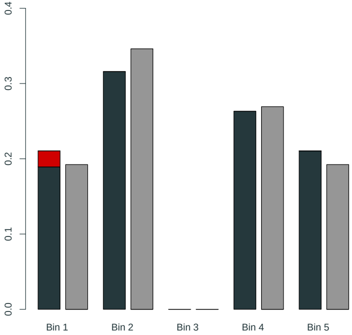


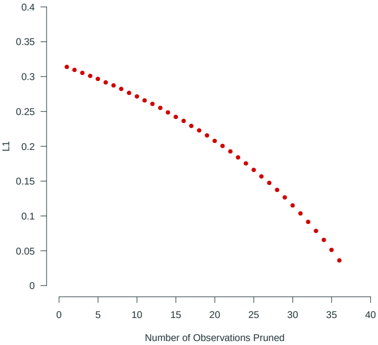


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

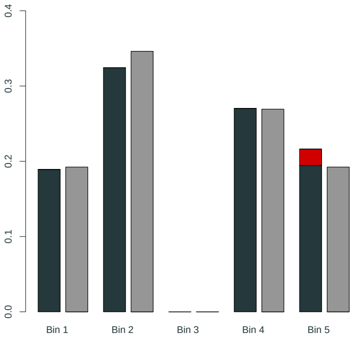


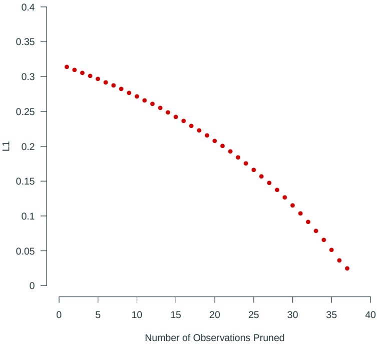


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

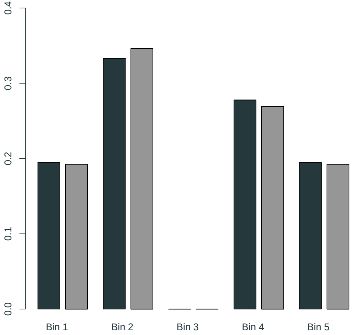


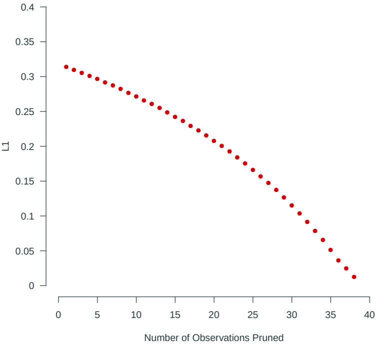


Constructing the L1/L2 SATT Frontier

Short version:

1. Calculate bins
2. Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units





Conclusions about Matching

Conclusions about Matching

- Matching:

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate
- In applications, focus on the substance:

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate
- In applications, focus on the substance:
  - All observational methods must assume ignorability

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate
- In applications, focus on the substance:
  - All observational methods must assume ignorability
  - Automated methods to choose variables: Insufficient

Conclusions about Matching

- Matching:
    - An excellent method of improving causal inferences
    - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate
- In applications, focus on the substance:
    - All observational methods must assume ignorability
    - Automated methods to choose variables: Insufficient
    - Claims of invariance: Always fail

Conclusions about Matching

- Matching:
  - An excellent method of improving causal inferences
  - Helps reduce: imbalance, model dependence, researcher discretion, bias
- Propensity Score Matching: low standards, dangerous in practice; DNR
- Most other methods of matching: excellent
- Coarsened Exact Matching: especially powerful; simple to understand and use
- Matching Frontier: Automates what's left to automate
- In applications, focus on the substance:
  - All observational methods must assume ignorability
  - Automated methods to choose variables: Insufficient
  - Claims of invariance: Always fail
  - Trying to be invariant to the substance: be wary of methods claiming to be invariant to what you know!

For more information, papers, & software

GaryKing.org

Part 4 (of 3), :-)

Matching Theories of Inference (in one slide)

Assumptions to Justify Current Practice

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T = 1|A) < 1$

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T = 1|A) < 1$
- Fits all common matching methods & practices; no asymptotics

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T = 1|A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; A too wide, n too small