Simplifying Matching Methods for Causal Inference

Gary King¹

Institute for Quantitative Social Science
Harvard University

(Talk at Stanford University, Department of Political Science, 1/14/2015)

¹GaryKing.org
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory: “A Theory of Statistical Inference for Matching Methods in Applied Causal Research” (Stefano Iacus, Gary King, Giuseppe Porro)

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

• Current practice, matching as preprocessing:
3 Problems, 3 Solutions

- Current practice, matching as preprocessing: violates current statistical theory.

- The most popular method (propensity score matching, used in 49,600 articles!) sounds magical

- Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:

“A Theory of Statistical Inference for Matching Methods in Applied Causal Research” (Stefano Iacus, Gary King, Giuseppe Porro)

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical

“Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously

“The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (Gary King, Christopher Lucas and Richard Nielsen)
3 Problems, 3 Solutions

- Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  \[ \rightsquigarrow \] “A Theory of Statistical Inference for Matching Methods in Applied Causal Research”
  (Stefano Iacus, Gary King, Giuseppe Porro)
3 Problems, 3 Solutions

- Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  \[ \rightsquigarrow \] “A Theory of Statistical Inference for Matching Methods in Applied Causal Research”
  (Stefano Iacus, Gary King, Giuseppe Porro)

- The most popular method (propensity score matching, used in 49,600 articles!) sounds magical
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  \[
  \leadsto \text{“A Theory of Statistical Inference for Matching Methods in Applied Causal Research”}
  \text{(Stefano Iacus, Gary King, Giuseppe Porro)}
  \]

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical
  \[
  \leadsto \text{“Why Propensity Scores Should Not Be Used for Matching”} \text{ (Gary King, Richard Nielsen)}
  \]
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  (Stefano Iacus, Gary King, Giuseppe Porro)

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical
  ↞ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
    (Stefano Iacus, Gary King, Giuseppe Porro)

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical
  ⇝ “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen)

• Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously
  ⇝ “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (Gary King, Christopher Lucas and Richard Nielsen)
Matching to Reduce Model Dependence
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)
Matching to Reduce Model Dependence

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

![Graph showing correlation between education (years) and outcome values.](image)
Matching to Reduce Model Dependence
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

![Graph showing scatter plot of Education (years) vs Outcome]

- **Education (years)**: 12, 14, 16, 18, 20, 22, 24, 26, 28
- **Outcome**: 0, 2, 4, 6, 8, 10, 12
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Current Practice: Matching as Preprocessing

- **Y_i** dep var, \( T_i \) (1 = treated, 0 = control), \( X_i \) confounders

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

- **Quantities of Interest:**
  1. **SATT**: Sample Average Treatment effect on the Treated: 
     \[
     SATT = \text{mean}_{i \in \{T_i = 1\}}(TE_i)
     \]
  2. **FSATT**: Feasible Average Treatment effect on the Treated

- **Estimate** \( Y_i(0) \) with \( Y_j \) from matched \( (X_i \approx X_j) \) control

- **Prune nonmatches**: reduces imbalance & model dependence

- **Big convenience**: Follow preprocessing with whatever statistical method you'd have used without matching
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
Current Practice: Matching as Preprocessing

- $Y_i$: dep var, $T_i$: (1=treated, 0=control), $X_i$: confounders
- Treatment Effect for treated observation $i$:

$$\text{TE}_i = Y_i(1) - Y_i(0)$$
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  \text{TE}_i = Y_i(1) - Y_i(0) = \text{observed} - \text{unobserved}
  \]
Current Practice: Matching as Preprocessing

- \( Y_i \) dep var, \( T_i \) (1=treated, 0=control), \( X_i \) confounders
- Treatment Effect for treated observation \( i \):
  \[
  \text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

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

- Quantities of Interest:

Prune nonmatches: reduces imbalance & model dependence

Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{mean}_{i \in \{T_i=1\}} (TE_i)$$
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  \text{TE}_i = Y_i - Y_i(0)
  = \text{observed} - \text{unobserved}
  \]

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{mean}_{i \in \{ T_i=1 \}} (\text{TE}_i)
     \]
  2. FSATT: Feasible Average Treatment effect on the Treated
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  \text{TE}_i = Y_i - Y_i(0)
  \]
  = observed – \textit{unobserved}

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{mean}_{i \in \{T_i=1\}} (\text{TE}_i)
     \]
  2. FSATT: Feasible Average Treatment effect on the Treated
- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) control
Current Practice: Matching as Preprocessing

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:

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

= observed − unobserved

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{mean}_{i \in \{T_i=1\}} (TE_i)$$

  2. FSATT: Feasible Average Treatment effect on the Treated

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) control
- Prune nonmatches: reduces imbalance & model dependence
Current Practice: Matching as Preprocessing

- \( Y_i \) dep var, \( T_i \) (1=treated, 0=control), \( X_i \) confounders
- Treatment Effect for treated observation \( i \):

\[
TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
\]

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

\[
\text{SATT} = \text{mean}_{i \in \{T_i=1\}} (TE_i)
\]

  2. FSATT: Feasible Average Treatment effect on the Treated

- Estimate \( Y_i(0) \) with \( Y_j \) from matched \((X_i \approx X_j)\) control
- Prune nonmatches: reduces imbalance & model dependence
- **Big convenience**: Follow preprocessing with whatever statistical method you’d have used without matching
Assumptions to Justify Current Practice

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

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

Alternatives Theory of Inference: It’s Gonna be OK!

• Assumes 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: \( \Pr(T=1 \mid A) < 1 \) (\( T=0,1 \) are both possible)

• Fits all common matching methods & practices; no asymptotics
• Easy extensions for: multi-level, continuous, & mismeasured treatments; \( A \) too wide, \( n \) too small
Assumptions to Justify Current Practice

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

- Assumes simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) | X$ (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!

- Assumes 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: $T \perp Y(0) | A$
  - 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
Assumptions to Justify Current Practice

**Existing Theory of Inference:** Stop What You’re Doing!

- Assumes simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - **Unconfoundedness:** \( T \perp Y (0) | X \) (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!

- Assumes 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:** \( T \perp Y (0) | A \)
  - **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
Assumptions to Justify Current Practice

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

- Assumes simple random sampling from a population

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

- Assumes stratified random sampling from a population
Assumptions to Justify Current Practice

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

- Assumes 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: Set-wide unconfoundedness
- Set-wide common support
Assumptions to Justify Current Practice

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

• Assumes 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!

• Assumes 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: $T \perp Y(0) | A$
  • Set-wide Common support: $\Pr(T = 1 | A) < 1$ ($T = 0, 1$ are both possible)
  • Fits all common matching methods & practices; no asymptotics
  • Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

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

- Assumes 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!

- Assumes 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: $\Pr(T = 1 \mid A) < 1$ (T = 0, 1 are both possible)
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

**Existing Theory of Inference: Stop What You’re Doing!**

- Assumes *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*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ are both possible)

**Alternative Theory of Inference: It’s Gonna be OK!**

- Assumes *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*: $\Pr(T = 1 \mid A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Assumptions to Justify Current Practice

**Existing Theory of Inference: Stop What You’re Doing!**

- Assumes *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*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!**

- Assumes 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*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!**

- Assumes stratified random sampling from a population
Assumptions to Justify Current Practice

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

- Assumes simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - Unconfoundedness: $T \perp Y(0) | X$ (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!

- Assumes 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!

• Assumes 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: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!

• Assumes 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!

- Assumes 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: \( \Pr(T = 1 \mid X) < 1 \) (\( T = 0, 1 \) 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!

- Assumes 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!**

- Assumes **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*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!**

- Assumes **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!

- Assumes 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: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!

- Assumes 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: $\Pr(T = 1 \mid A) < 1$
Assumptions to Justify Current Practice

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

- Assumes **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*: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!

- Assumes **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*: $\Pr(T = 1 \mid A) < 1$
- Fits all common matching methods & practices; no asymptotics
Assumptions to Justify Current Practice

**Existing Theory of Inference: Stop What You’re Doing!**

- Assumes 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: $\Pr(T = 1 \mid X) < 1$ ($T = 0, 1$ 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!**

- Assumes 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: $\Pr(T = 1 \mid A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; $A$ too wide, $n$ too small
Approximating Randomized Experiments

Types of experiments:

1. Compete Randomization: Treatment assignment by coin flips
   - Balance on $X$: only on average
   - Balance on unmeasured vars: only on average

2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins
   - Balance on $X$: perfect in sample
   - Balance on unmeasured vars: only on average

Fully blocked dominates complete randomization for:
imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

Matching methods approximate which experiment?

PSM: complete randomization
Other methods: fully blocked

As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

• Types of experiments:

1. Compete Randomization: Treatment assignment by coin flips
   ⇝ Balance on $X$: only on average
   ⇝ Balance on unmeasured vars: only on average

2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins
   ⇝ Balance on $X$: perfect in sample
   ⇝ Balance on unmeasured vars: only on average

• Fully blocked dominates complete randomization for:
  imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

• Matching methods approximate which experiment?
  • PSM: complete randomization
  • Other methods: fully blocked

⇒ As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

- Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips

Other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     \[ \leadsto \text{Balance on} \ X: \text{only on average} \]
Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for:
  - imbalance,
  - model dependence,
  - power,
  - efficiency,
  - bias,
  - research costs,
  - robustness.

- Matching methods approximate which experiment?
  - PSM: complete randomization
  - Other methods: fully blocked

  As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins

Matching methods approximate which experiment?
- PSM: complete randomization
- Other methods: fully blocked

In practice, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

• Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: **perfect in sample**

• Matching methods approximate which experiment?
  - PSM: compete randomization
  - Other methods: fully blocked

• Fully blocked dominates compete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

• As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

- Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

Matching methods approximate which experiment?

- PSM: complete randomization
- Other methods: fully blocked

As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- **Fully blocked dominates complete randomization**
Approximating Randomized Experiments

• Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

• Fully blocked dominates complete randomization for:
Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for:
  imbalance,
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for: imbalance, model dependence,
Approximating Randomized Experiments

- Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for:
  imbalance, model dependence, power,
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for:
  imbalance, model dependence, power, efficiency,
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- **Fully blocked** dominates complete randomization for: imbalance, model dependence, power, efficiency, bias,
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs,
Approximating Randomized Experiments

- Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
Approximating Randomized Experiments

• Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     ⇝ Balance on $X$: only on average
     ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins
     ⇝ Balance on $X$: perfect in sample
     ⇝ Balance on unmeasured vars: only on average

• Fully blocked dominates complete randomization for:
  imbalance, model dependence, power, efficiency, bias,
  research costs, and robustness.

• Matching methods approximate which experiment?
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- **Fully blocked** dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

- **Matching methods approximate which experiment?**
  - PSM: complete randomization
Approximating Randomized Experiments

- **Types of experiments:**
  1. **Compete Randomization:** Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. **Fully Blocked:** Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- **Fully blocked dominates complete randomization** for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

- **Matching methods approximate which experiment?**
  - PSM: complete randomization
  - Other methods: fully blocked
Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - Balance on $X$: only on average
     - Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on $X$ (exactly), then flip coins
     - Balance on $X$: perfect in sample
     - Balance on unmeasured vars: only on average

- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

- Matching methods approximate which experiment?
  - PSM: complete randomization
  - Other methods: fully blocked

- As we show, other methods usually dominate PSM
Approximating Randomized Experiments

• Types of experiments:
  1. **Compete Randomization**: Treatment assignment by coin flips
     ⇝ Balance on $X$: only on average
     ⇝ Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on $X$ (exactly), then flip coins
     ⇝ Balance on $X$: perfect in sample
     ⇝ Balance on unmeasured vars: only on average

• Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

• Matching methods approximate which experiment?
  • PSM: complete randomization
  • Other methods: fully blocked

• As we show, other methods usually dominate PSM
  (but wait, it gets worse for PSM)
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, . . .
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - \( \text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Distance \( X_i, X_j \) = \( \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching  
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)  
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$  
   - Match each treated unit to the nearest control unit

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance\((X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}\)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, . . .
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance($X_i, X_j$) = $\sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $> caliper$

2. Estimation
   Difference in means or a model

3. Checking
   Measure imbalance, tweak, repeat, ...
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Distance\( (X_i, X_j) = \sqrt{(X_i - X_j)'S^{-1}(X_i - X_j)} \)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > *caliper*

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, ...
Mahalanobis Distance Matching

Education (years)

<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>28</td>
</tr>
</tbody>
</table>

Age vs. Education (years)
Mahalanobis Distance Matching

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

Education (years)
Mahalanobis Distance Matching

![Graph showing Mahalanobis Distance Matching with Education (years) on the x-axis and Age on the y-axis. The graph includes data points and connecting lines.]
Method 2: Coarsened Exact Matching

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

2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds

3. Checking Determine matched sample size, tweak, repeat, . . .
   - Easier, but still iterative
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation
   • Difference in means or a model
     Need to weight controls in each stratum to equal treateds

3. Checking
   • Determine matched sample size, tweak, repeat, . . .
     Easier, but still iterative
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess (Matching)**
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units

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

3. **Checking** Determine matched sample size, tweak, repeat, ...
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

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

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
   - Need to weight controls in each stratum to equal treateds

3. **Checking** Determine matched sample size, tweak, repeat, . . .
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation**  Difference in means or a model
   - Need to weight controls in each stratum to equal treateds

3. **Checking**  Determine matched sample size, tweak, repeat, . . .
   - Easier, but still iterative
Coarsened Exact Matching
Coarsened Exact Matching

Education vs. Age

Education

12 14 16 18 20 22 24 26 28

Age

20
30
40
50
60
70
80

CCC C
CC CC
C CC C CCC CCCC CCC CC CCC CCCCCC
C CCC CC C

T T
T T
TT TT T TT
TTT TT
T
T TT
Coarsened Exact Matching

Education

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

Drinking age

HS BA MA PhD 2nd PhD

Senior Discounts

Retirement

Old

Drinking age

Education

Don't trust anyone over 30

The Big 40

Senior Discounts

Retirement

Old

Drinking age
Method 3: Propensity Score Matching

1. Preprocess (Matching)
   • Reduce $k$ elements of $X$ to scalar
     \[
     \pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}
     \]
   • Distance($X_i$, $X_j$) = $|\pi_i - \pi_j|$
   • Match each treated unit to the nearest control unit
   • Control units: not reused; pruned if unused
   • Prune matches if Distance $> \text{caliper}$

2. Estimation Difference in means or a model

3. Checking Measure imbalance, tweak, repeat, . . .
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar $\pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}$
   - Distance($X_i$, $X_j$) = $|\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > caliper

2. Estimation
   - Difference in means or a model

3. Checking
   - Measure imbalance, tweak, repeat, ...
Method 3: Propensity Score Matching  
(Approximates Completely Randomized Experiment)

1. **Preprocess** (Matching)

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

3. **Checking** Measure imbalance, tweak, repeat, ...
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
   \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-x_i\beta}} \]

2. **Estimation**  Difference in means or a model
3. **Checking**  Measure imbalance, tweak, repeat, ...
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-x_i\beta}} \]
   - Distance($X_i, X_j$) = $|\pi_i - \pi_j|$

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   • Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}} \]
   • Distance($X_i, X_j$) = $|\pi_i - \pi_j|$
   • Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar
     $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance($X_i, X_j$) = $|\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce \( k \) elements of \( X \) to scalar
     \[
     \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}
     \]
   - Distance \((X_i, X_j) = |\pi_i - \pi_j|\)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance > *caliper*

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
     $$\pi_i \equiv \Pr(T_i = 1 | X) = \frac{1}{1 + e^{-X_i \beta}}$$
   - Distance($X_i, X_j$) = $|\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance $> \text{caliper}$

2. **Estimation** Difference in means or a model
3. **Checking** Measure imbalance, tweak, repeat, . . .
Propensity Score Matching

![Graph showing the relationship between age and education years. The graph displays scattered points representing different data points, with some labeled 'C' and others 'T'.]
Propensity Score Matching

Education (years) vs. Age

Age

Education (years)

Propensity Score
Propensity Score Matching

![Graph showing the relationship between Education (years) and Age with points marked by 'C' and 'T'.]
Consequences of Matching with PSM

- Lots of information left on the table
- Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway

The Curse of Dimensionality

The Promise: avoid it by balancing on $\pi$ rather than $X$

The Reality: the paradox is worse with more covariates
Consequences of Matching with PSM

- Lots of information left on the table

- Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)

- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- The Curse of Dimensionality
  - The Promise: avoid it by balancing on $\pi$ rather than $X$
  - The Reality: the paradox is worse with more covariates
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
- The information loss is not innocuous:
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)
- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway
- The Curse of Dimensionality
Consequences of Matching with PSM

- Lots of information left on the table
  - Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)

- The information loss is not innocuous:
  - If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- The Curse of Dimensionality
  - The Promise: avoid it by balancing on $\pi$ rather than $X$
Consequences of Matching with PSM

• Lots of information left on the table
  • Full blocking can greatly increase efficiency (Imai, King, and Nall: up to 600% difference in SEs)

• The information loss is not innocuous:
  • If data are balanced to begin with, or after some pruning, the paradox will kick in and imbalance will get worse
  • If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

• The Curse of Dimensionality
  • The Promise: avoid it by balancing on $\pi$ rather than $X$
  • The Reality: the paradox is worse with more covariates
PSM is Blind Where Others Can See
PSM is Blind Where Others Can See
The Propensity Score Paradox

Finkle et al. (2012)

Nielsen et al. (2011)
What Does PSM Match?

MDM Matches

PSM Matches

Controls: \( X_1, X_2 \sim \text{Uniform}(0,5) \)

Treateds: \( X_1, X_2 \sim \text{Uniform}(1,6) \)
PSM Increases Model Dependence

\[ Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i \]
\[ \epsilon_i \sim N(0, 1) \]
The Matching Frontier

- Bias-Variance trade off
  \[ \Rightarrow \]
  Imbalance-Trade Off

- Frontier = matched dataset with lowest imbalance for each \( n \)

- (Maybe we can beat MDM/CEM for a given #pruned?)

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

- Result:
  - Simple to use
  - 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
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
The Matching Frontier

• Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
• (Maybe we can beat MDM/CEM for a given #pruned?)
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

- Simple to use
- 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
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
The Matching Frontier

- Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each $n$
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
The Matching Frontier

- Bias-Variance trade off → Imbalance-\(n\) Trade Off
  Frontier = matched dataset with lowest imbalance for each \(n\)
- (Maybe we can beat MDM/CEM for a given \#pruned?)
- 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

- Bias-Variance trade off $\leadsto$ Imbalance-$n$ Trade Off Frontier = matched dataset with lowest imbalance for each $n$
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Simple to use
The Matching Frontier

• Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
• (Maybe we can beat MDM/CEM for a given #pruned?)
• To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
• Result:
  • Simple to use
  • No need to choose or use a matching method
The Matching Frontier

• Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
• (Maybe we can beat MDM/CEM for a given #pruned?)
• To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
• Result:
  • Simple to use
  • No need to choose or use a matching method
  • All solutions are optimal
The Matching Frontier

• Bias-Variance trade off $\sim$ Imbalance-$n$ Trade Off
  Frontier = matched dataset with lowest imbalance for each $n$
• (Maybe we can beat MDM/CEM for a given #pruned?)
• To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

• Result:
  • Simple to use
  • No need to choose or use a matching method
  • All solutions are optimal
  • No iteration or diagnostics required
The Matching Frontier

- Bias-Variance trade off \(\sim\) Imbalance-\(n\) Trade Off Frontier = matched dataset with lowest imbalance for each \(n\)
- (Maybe we can beat MDM/CEM for a given \#pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching

- Result:
  - Simple to use
  - 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
How hard is the frontier to calculate?

• Consider 1 point on the SATT frontier:

\[ \text{Starting with matrix of } N \text{ control units } X_0 \]

\[ \text{Calculate imbalance for all } \binom{N}{n} \text{ subsets of rows of } X_0 \]

\[ \text{Choose subset with lowest imbalance} \]

\[ \text{Evaluations needed to compute the entire frontier:} \]

\[ \binom{N}{n} \text{ evaluations for each sample size } n = N, N - 1, \ldots, 1 \]

\[ \text{The combination is the (gargantuan) "power set"} \]

\[ \text{e.g., } N > 300 \text{ requires more imbalance evaluations than} \]

\[ \text{elementary particles in the universe} \]

\[ \Rightarrow \text{It's hard to calculate!} \]

• We develop algorithms for the (optimal) frontier which:

\[ \text{run very fast} \]

\[ \text{operate as "greedy" but we prove are optimal} \]

\[ \text{do not require evaluating every subset} \]

\[ \text{work with very large data sets} \]

\[ \text{is the exact frontier (no approximation or estimation)} \]

\[ \Rightarrow \text{It's easy to calculate!} \]
How hard is the frontier to calculate?

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

• 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

  $\Rightarrow$ 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)

  $\Rightarrow$ It's easy 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 $X_0$

\[\binom{N}{n}\] evaluations for each 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

\[ \binom{N}{n} \] evaluations for each sample size $n = N, N-1, ..., 1$

The combination is the (gargantuan) “power set”

- e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

\[ \Rightarrow \text{It's hard to calculate!} \]

- 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)

\[ \Rightarrow \text{It's easy to calculate!} \]
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

- ⇝ 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)

- ⇝ It's easy 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 $n = N, N - 1, \ldots, 1$

$\Rightarrow$ 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)

$\Rightarrow$ It's easy 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”
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

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)

It's easy 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!
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
  - \( \Rightarrow \) 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 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
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
  • \( \leadsto \) 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 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
  - $\Rightarrow$ 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 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
  - $\Rightarrow$ 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
  - $\Rightarrow$ 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
  - $\Rightarrow$ 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)
  - $\Rightarrow$ It’s easy to calculate!
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

Remaining Data

Frontier

Covariate 1

Covariate 2

-1.0
-0.5
0.0
0.5
1.0

0
5
10
15
20

0.0
0.1
0.2
0.3
0.4

Treated
Control
Next to remove

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

**Remaining Data**

- Covariate 1
- Covariate 2

**Frontier**

- Number of Observations Dropped
- Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2

-1.0  -0.5  0.0  0.5  1.0

•
•
•
•
•
•

Treated
Control
Next to remove

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy

0  0.5  1.0  1.5  2.0

0.0
0.1
0.2
0.3
0.4

Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Treated
Control
Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Covariate 1

Covariate 2

Treated
Control
Next to remove

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

Treated
Control
Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy

Covariate 1

Covariate 2

-1.0 -0.5 0.0 0.5 1.0

-1.0 -0.5 0.0 0.5 1.0

0.0

0.1

0.2

0.3

0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

Number of Observations Dropped: 0, 5, 10, 15, 20

Average Mahalanobis Discrepancy: 0.0, 0.1, 0.2, 0.3, 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

Covariate 1
Covariate 2

Number of Observations Dropped
Average Mahalanobis Discrepancy

0 5 10 15 20
0.0 0.1 0.2 0.3 0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Covariates 1 and 2 are plotted on the left side of the graph.
- Treated, Control, and Next to remove are indicated by different markers.
- The average Mahalanobis discrepancy is plotted on the right side.
- The number of observations dropped is shown on the x-axis.

Graphical representation and data visualization.
Constructing the FSATT Mahalanobis Frontier

### Remaining Data

- **Covariate 1**
  - Values: $-1.0, -0.5, 0.0, 0.5, 1.0$

- **Covariate 2**
  - Values: $-1.0, -0.5, 0.0, 0.5, 1.0$

- **Points:**
  - Red circles: Treated
  - Black circles: Control
  - Blue circle: Next to remove

### Frontier

- **Average Mahalanobis Discrepancy**
  - Values: $0.0, 0.1, 0.2, 0.3, 0.4$

- **Number of Observations Dropped**
  - Values: $0, 5, 10, 15, 20$

- **Graphs:**
  - Covariate 1 vs. Covariate 2
  - Average Mahalanobis Discrepancy vs. Number of Observations Dropped

- **Legend:**
  - Treated
  - Control
  - Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Average Mahalanobis Discrepancy
- Number of Observations Dropped

- 0.0
- 0.1
- 0.2
- 0.3
- 0.4

- 0
- 5
- 10
- 15
- 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2
- Treated
- Control
- Next to remove

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

Legend:
- Red circle: Treated
- Gray circle: Control
- Blue circle: Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- ● Treated
- ○ Control
- ○ Next to remove

Covariate 1
Covariate 2

Number of Observations Dropped
Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Covariate 1
Covariate 2

-1.0 -0.5 0.0 0.5 1.0
-1.0
-0.5
0.0
0.5
1.0
●
●
●
●
●

Treated
Control
Next to remove

Frontier

Number of Observations Dropped

Average Mahalanobis Discrepancy

0.0
0.1
0.2
0.3
0.4 ●
●
● ●
●
●
●
●
●
● ●
●
●
●
●

• Warning: figure omits details and the proof!
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Treated
- Control
- Next to remove

- Warning: figure omits details and the proof!
- Very fast; works with any continuous imbalance metric
Constructing the L1/L2 SATT Frontier

![Graph showing frequency and number of observations dropped in different bins for treatment and control groups.](image-url)
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distribution across different bins.]

- **Bin 1**: Treatment 5, Control 4
- **Bin 2**: Treatment 6, Control 7
- **Bin 3**: Treatment 2, Control 3
- **Bin 4**: Treatment 3, Control 6
- **Bin 5**: Treatment 3, Control 2
- **Bin 6**: Treatment 1, Control 2

![Graph showing L2 vs. number of observations dropped.]

- Number of observations dropped along the x-axis.
- L2 values along the y-axis.

L2 values: 0.00, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12
Constructing the L1/L2 SATT Frontier

![Graph showing frequency and L2 values for different bins.]

- **Bins:** Bin 1, Bin 2, Bin 3, Bin 4, Bin 5, Bin 6
- **Frequency Counts:**
  - Bin 1: Treatment 5, Control 4
  - Bin 2: Treatment 6, Control 7
  - Bin 3: Treatment 2, Control 3
  - Bin 4: Treatment 3, Control 5
  - Bin 5: Treatment 3, Control 2
  - Bin 6: Treatment 1, Control 2

- **L2 Values:**
  - L2 values range from 0.00 to 0.12
  - L2 values are plotted against the number of observations dropped, which ranges from 0 to 10.
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency in different bins for Treatment and Control groups.](chart)

Frequency

- Bin1: 5 for Treatment, 4 for Control
- Bin2: 6 for Treatment, 7 for Control
- Bin3: 2 for Treatment, 3 for Control
- Bin4: 3 for Treatment, 4 for Control
- Bin5: 3 for Treatment, 2 for Control
- Bin6: 1 for Treatment, 2 for Control

![Graph showing number of observations dropped vs. L2.](graph)

- Number of Observations Dropped: 0, 2, 4, 6, 8, 10
- L2 values: 0.00, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12

- Each point on the graph corresponds to a decrease in the number of observations dropped with a corresponding decrease in L2.
Constructing the L1/L2 SATT Frontier

![Bar Chart and Line Graph Illustrating Frequency and L2 Values Across Bins]

- **Bar Chart**: Frequency distribution across different bins (Bin1 to Bin6) for Treatment and Control groups.
- **Line Graph**: L2 values plotted against the number of observations dropped.

**Key Points**:
- **Number of Observations Dropped**: Frequency counts for each bin.
- **L2 Values**: Decrease in L2 values with an increase in the number of observations dropped.

---

**Legend**:
- Blue: Treatment
- Red: Control
Constructing the L1/L2 SATT Frontier

Number of Observations Dropped

L2

0.00 0.02 0.04 0.06 0.08 0.10 0.12

0 2 4 6 8 10

Number of Observations Dropped

Frequency

Bin1  Bin2  Bin3  Bin4  Bin5  Bin6

4    6    3    3    2    1

4    6    2    4    3    2

Treatment □ Control

L2
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency distribution across bins]

- Bin 1: Treatment 4, Control 2
- Bin 2: Treatment 6, Control 6
- Bin 3: Treatment 2, Control 2
- Bin 4: Treatment 4, Control 3
- Bin 5: Treatment 3, Control 2
- Bin 6: Treatment 1, Control 2

![Graph showing L2 vs. number of observations dropped]

- L2 values decrease as the number of observations dropped increases.

Number of Observations Dropped
Constructing the L1/L2 SATT Frontier

![Chart showing frequency distribution and L2 values.](chart.png)
Constructing the L1/L2 SATT Frontier

![Graph showing frequency and L2 values across different bins.](image)
Constructing the L1/L2 SATT Frontier

![Graph showing frequency distribution and L2 values](image-url)
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
- Works very fast, even with very large data sets
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence

• Some mistakes with PSM:
  • Controlling for irrelevant covariates;
  • Adjusting experimental data;
  • Reestimating propensity score after eliminating noncommon support;
  • 1/4 caliper on propensity score;
  • Not switching to other methods.

• Theory of Inference for Matching
  • Switch from simple to stratified random sampling
  • Justifies current practices
  • Clarifies how to improve inferences

• Using more information is simpler and more powerful
Conclusions

- The Matching Frontier

- Fast; easy; no iteration; Software: MatchingFrontier
- No need to choose among matching methods
- Optimal results from your choice of imbalance metric

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence

- Some mistakes with PSM:
  - Controlling for irrelevant covariates;
  - Adjusting experimental data;
  - Reestimating propensity score after eliminating non-common support;
  - 1/4 caliper on propensity score;
  - Not switching to other methods.

- Theory of Inference for Matching
  - Switch from simple to stratified random sampling
  - Justifies current practices
  - Clarifies how to improve inferences

- Using more information is simpler and more powerful
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Propensity score matching:
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
Conclusions

• **The Matching Frontier**
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• **Propensity score matching:**
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM:
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates;
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data;
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support;
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score;
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- **Theory of Inference for Matching**
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- **Theory of Inference for Matching**
  - Switch from simple to stratified random sampling
Conclusions

• **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

• **Propensity score matching**:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating non-common support; 1/4 caliper on propensity score; Not switching to other methods.

• **Theory of Inference for Matching**
  - Switch from simple to stratified random sampling
  - Justifies current practices
Conclusions

- **The Matching Frontier**
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

- **Propensity score matching:**
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- **Theory of Inference for Matching**
  - Switch from simple to stratified random sampling
  - Justifies current practices
  - Clarifies how to improve inferences
Conclusions

• The Matching Frontier
  • Fast; easy; no iteration; Software: MatchingFrontier
  • No need to choose among matching methods
  • Optimal results from your choice of imbalance metric

• Propensity score matching:
  • Approximates complete, not fully blocked, experiments
  • Ignores information; exacerbates model dependence
  • Some mistakes with PSM: Controlling for irrelevant covariates;
    Adjusting experimental data; Reestimating propensity score
    after eliminating noncommon support; 1/4 caliper on
    propensity score; Not switching to other methods.

• Theory of Inference for Matching
  • Switch from simple to stratified random sampling
  • Justifies current practices
  • Clarifies how to improve inferences

• Using more information is simpler and more powerful
For more information, papers, & software

GaryKing.org