Simplifying Matching Methods for Causal Inference

Gary King

Institute for Quantitative Social Science
Harvard University

(Talk at MIT, Political Methodology Series, 3/16/2015)
3 Problems, 3 Solutions
3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 
  - Violates current statistical theory. 
  - 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.
3 Problems, 3 Solutions

• Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
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)
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:
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)
3 Problems, 3 Solutions

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

• The most popular method (propensity score matching, used in 49,600 articles!) sounds magical:
  \[ \rightsquigarrow \text{“Why Propensity Scores Should Not Be Used for Matching”} \]
  (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:
  \[ \Rightarrow \text{ “A Theory of Statistical Inference for Matching Methods in Applied Causal Research”} \]
  (Stefano Iacus, Gary King, Giuseppe Porro)

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

- Matching methods optimize either imbalance (≈ bias) or # units pruned (≈ variance); users need both simultaneously’:
  \[ \Rightarrow \text{ “The Balance-Sample Size Frontier in Matching Methods for Causal Inference”} \]
  (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*)
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)
The Advantage of Matching

Without Matching:
The Advantage of Matching

Without Matching:

Imbalance
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Advantage of Matching

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
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 $i$: $TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$

- 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
- 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 \( i \):

\[
TE_i = Y_i(1) - Y_i(0) = \text{observed} - \text{unobserved}
\]
Current Practice: Matching as Preprocessing

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

$$TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$
Current Practice: Matching as Preprocessing

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

- Quantities of Interest:
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:

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

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

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

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

- 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
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

• 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: \( \Pr(T = 1 \mid X) < 1 \) for both \( T = 0 \) and \( T = 1 \)

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: \( \Pr(T = 1 \mid A) < 1 \)
  • 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!

- Framework: 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!
- 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: $T \perp Y(0) | A$
  - 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!

**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!

- Framework: stratified random sampling from a population

\[ T \perp Y(0) | X \] (Healthy & unhealthy get meds)

\[ \text{Pr}(T = 1 | X) < 1 \] (T = 0, 1 are both possible)

Set-wide Unconfoundedness:

\[ T \perp Y(0) | A \]

Set-wide Common support:

\[ \text{Pr}(T = 1 | A) < 1 \]

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!

- 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!

- 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: $T \perp Y(0) | A$
- 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!

- 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!

- 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: $T \perp Y(0) | A$
  - 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!

- 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!

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

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

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*: \( \Pr(T = 1 \mid A) < 1 \)
  - 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!

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

- 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**: $\Pr(T = 1 \mid A) < 1$
  - 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!

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

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

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

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

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

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

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

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

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

- Framework: 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!

- 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: $T \perp Y(0) | A$
  - 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
Approximating Randomized 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:
  - Compete Randomization: Treatment assignment by coin flips
    - Balance on $X$: only on average
    - Balance on unmeasured vars: only on average
  - 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

- 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
  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
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

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
     \(\leadsto\) Balance on \(X\): only on average
     \(\leadsto\) Balance on unmeasured vars: only on average
  2. **Fully Blocked**: Match pairs on \(X\) (exactly), then flip coins
     \(\leadsto\) Balance on \(X\): **perfect in sample**
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)
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 $> \text{caliper}$

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

Age

Education (years)
Mahalanobis Distance Matching

Age

Education (years)
Mahalanobis Distance Matching

Age

Education (years)

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

C

C

CC

C

C

C

C

C

CC

C

CC

C

CC

C

C

T TTT T

T

T

T

T

T

T

T

TT

T

T

T

T

T

C

C

C

C

C

C

C

C

T

T

T

T

T

T

T

T

T

T

T

T

T
Mahalanobis Distance Matching

Education (years)

Age

Education (years)
Mahalanobis Distance Matching

Education (years)

Age

12  14  16  18  20  22  24  26  28
20
30
40
50
60
70
80
T  T
T  T
TT  TT  T  TT
TTT  TT
T
T  TT
C
C
C  C
CC
C
C
C  CCC
C
C
CC  C  C  C
C
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 $X$
   • 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

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, …
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

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

Education:
- HS
- BA
- MA
- PhD
- 2nd PhD

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

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

Education:
- HS
- BA
- MA
- PhD
- 2nd PhD

Don't trust anyone over 30
Coarsened Exact Matching

Education

HS  BA  MA  PhD  2nd PhD

Don't trust anyone over 30
The Big 40
Senior Discounts
Retirement
Old
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 > 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 $>$ *caliper*

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

Age

Education (years)

Propensity Score
Propensity Score Matching

Education (years) | Age
---|---
12 | 16 | 20 | 24 | 28 | 30 | 40 | 50 | 60 | 70 | 80

Propensity Score

1
0

Propensity Score
Propensity Score Matching

Age

Education (years)

Propensity Score
Propensity Score Matching

Education (years)

Age

12 16 20 24 28

20

30

40

50

60

70

80

1

0

Propensity Score

C

CC

C

CC

C

C

CC

C

CC

CC

T

TTT

T

TT

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T

T
Propensity Score Matching

Education (years) vs. Age

- Education values: 12, 16, 20, 24, 28
- Age values: 20, 30, 40, 50, 60, 70, 80
- Symbols: C, T
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata)
  - Matching is at random

- **Random matching increases imbalance:**
  - Approximating complete randomization (by pruning) \( \Rightarrow \) higher imbalance \( \Rightarrow \) more inefficiency

- If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance \( \Rightarrow \) Inefficiency \( \Rightarrow \) Model dependence \( \Rightarrow \) Bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on \( \pi \) rather than \( X \)
  - The Reality: The PSM Paradox is bigger with more covariates
PSM’s Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
  - The PSM Paradox: If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \Rightarrow \) matching is at random
    - Random matching increases imbalance!
    - Approximating complete randomization (by pruning) \( \Rightarrow \) higher imbalance \( \Rightarrow \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- PSM is Biased:
  - Imbalance \( \Rightarrow \) Inefficiency \( \Rightarrow \) Model dependence \( \Rightarrow \) Bias

- Curse of Dimensionality Problems:
  - The Promise: avoid it by balancing on \( \pi \) rather than \( X \)
  - The Reality: The PSM Paradox is bigger with more covariates
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata)
  - Matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \Rightarrow \) higher imbalance \( \Rightarrow \) more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway

- **PSM is Biased:**
  - Imbalance \( \Rightarrow \) inefficiency \( \Rightarrow \) model dependence \( \Rightarrow \) bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on \( \pi \) rather than \( X \)
  - The Reality: The PSM Paradox is bigger with more covariates
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata)
  - Matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\Rightarrow$ higher imbalance $\Rightarrow$ more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance $\Rightarrow$ Inefficency $\Rightarrow$ Model dependence $\Rightarrow$ Bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on $\pi$ rather than $X$
  - The Reality: The PSM Paradox is bigger with more covariates
PSM’s Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- The PSM Paradox:
  \[ \hat{\pi} \approx 0.5 \text{ (or constant within strata)} \Rightarrow \text{matching is at random} \]
  
  - Random matching increases imbalance!
  
  \[ \Rightarrow \text{higher imbalance} \Rightarrow \text{more inefficiency} \]
  
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- PSM is Biased:
  \[ \text{Imbalance } \Rightarrow \text{Inefficiency } \Rightarrow \text{Model dependence } \Rightarrow \text{Bias} \]

- Curse of Dimensionality Problems:
  \[ \text{The Promise: avoid it by balancing on } \pi \text{ rather than } X \]
  
  \[ \text{The Reality: The PSM Paradox is bigger with more covariates} \]
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \Rightarrow \) matching is at random
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning,
    $\hat{\pi} \approx 0.5$ (or constant within strata) $\Rightarrow$ matching is at random
  - Random matching increases imbalance!

- **PSM is Biased:**
  - Imbalance $\Rightarrow$ Inefficency $\Rightarrow$ Model dependence $\Rightarrow$ Bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on $\pi$ rather than $X$
  - The Reality: The PSM Paradox is bigger with more covariates
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \rightsquigarrow \) higher imbalance \( \rightsquigarrow \) more inefficiency
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \Rightarrow \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \Rightarrow \) higher imbalance \( \Rightarrow \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\Rightarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\Rightarrow$ higher imbalance $\Rightarrow$ more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \leadsto \) higher imbalance \( \leadsto \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance \( \leadsto \) Inefficiency \( \leadsto \) Model dependence \( \leadsto \) Bias
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \implies \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \implies \) higher imbalance \( \implies \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance \( \implies \) Inefficency \( \implies \) Model dependence \( \implies \) Bias

- **Curse of Dimensionality Problems:**
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \Rightarrow \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \Rightarrow \) higher imbalance \( \Rightarrow \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance \( \Rightarrow \) Inefficiency \( \Rightarrow \) Model dependence \( \Rightarrow \) Bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on \( \pi \) rather than \( X \)
PSM’s Statistical Properties

- **PSM is Inefficient:**
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

- **The PSM Paradox:**
  - If data are balanced to begin with, or after some pruning, \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) \( \rightsquigarrow \) higher imbalance \( \rightsquigarrow \) more inefficiency
  - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway

- **PSM is Biased:**
  - Imbalance \( \rightsquigarrow \) Inefficiency \( \rightsquigarrow \) Model dependence \( \rightsquigarrow \) Bias

- **Curse of Dimensionality Problems:**
  - The Promise: avoid it by balancing on \( \pi \) rather than \( X \)
  - The Reality: The PSM Paradox is bigger with more covariates
PSM is Blind Where Other Methods Can See
PSM is Blind Where Other Methods Can See
PSM is Blind Where Other Methods Can See
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 & Bias

Model Dependence

Bias

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

Finkle et al. (2012)

Nielsen et al. (2011)
The Matching Frontier

• Bias-Variance trade off ⇝ Imbalance-Trade Off

Frontier = matched dataset with lowest imbalance for each 

• (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 \( \rightsquigarrow \) 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:
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
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 $\sim$ 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:
The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ 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:

- 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

Therefore, 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)

Therefore, 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 ${N \choose n}$ subsets of rows of $X_0$
  - Choose subset with lowest imbalance

  Evaluations needed to compute the entire frontier:
  
  - $\left( {N \choose n} \right)$ 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 $X_0$
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 $X_0$

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

$\Rightarrow$ 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:
  - 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 $X_0$
  • Choose subset with lowest imbalance

• Evaluations needed to compute the entire frontier:
  • $\binom{N}{n}$ evaluations for each sample size $n = N, N - 1, \ldots, 1$
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 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!
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
  - \( \sim \) 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
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

\[20/26\]
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

$\Rightarrow$ 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”
  • 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!
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

- **Covariate 1**
- **Covariate 2**

-1.0  -0.5  0.0  0.5  1.0

### Frontier

- **Treated**
- **Control**
- **Next to remove**

- **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

- Treated
- Control
- Next to remove

Number of Observations Dropped

Average Mahalanobis Discrepancy
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Treated
- Control
- Next to remove

Frontier

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

Remaining Data

- Treated
- Control
- Next to remove

Frontier

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

Remaining Data

- Covariate 1
- Covariate 2

Frontier

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

Legend:
- Treated
- Control
- Next to remove
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

Frontier

Treated
Control
Next to remove

Covariate 1
Covariate 2

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

Average Mahalanobis Discrepancy

Number of Observations Dropped
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

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

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:
- Treated
- Control
- Next to remove
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

Frontier

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

Number of Observations Dropped

Average Mahalanobis Discrepancy

0.0
0.1
0.2
0.3
0.4

0 5 10 15 20

0.0
0.1
0.2
0.3
0.4
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

-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

0 5 10 15 20

0.0

0.1

0.2

0.3

0.4

Average Mahalanobis Discrepancy

Number of Observations Dropped

Covariate 1

Covariate 2

Remaining Data

Covariate 1

Covariate 2
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

- Covariate 1
- Covariate 2

- Treated
- Control
- Next to remove

- Number of Observations Dropped
- Average Mahalanobis Discrepancy

- 0 5 10 15 20
- 0.0 0.1 0.2 0.3 0.4

- Covariate 1
- Covariate 2

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

- ●
- ●
- ●
- ●
- ●
- ●
- ●
- ●
- ●

- Next to remove
Constructing the FSATT Mahalanobis Frontier

Remaining Data

- Covariate 1
- Covariate 2

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

Average Mahalanobis Discrepancy
0 5 10 15 20
0.0
0.1
0.2
0.3
0.4 ●
●
●
●
●
Number of Observations Dropped
0 5 10 15 20
Constructing the FSATT Mahalanobis Frontier

Remaining Data

Frontier

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

Average Mahalanobis Discrepancy

Number of Observations Dropped

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

Remaining Data

Frontier

- 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 for different bins.]

- Bin 1: Frequency 5
- Bin 2: Frequency 6
- Bin 3: Frequency 7
- Bin 4: Frequency 7
- Bin 5: Frequency 3
- Bin 6: Frequency 2

Number of Observations Dropped:

- L2: 23 / 26
Constructing the L1/L2 SATT Frontier

[Bar chart showing frequency distribution for different bins labeled Bin1 to Bin6, with blue and red bars indicating Treatment and Control respectively.]

[Graph showing the number of observations dropped against L2 values.]
Constructing the L1/L2 SATT Frontier

<table>
<thead>
<tr>
<th>Bin</th>
<th>Frequency</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin1</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Bin2</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Bin3</td>
<td>7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Bin4</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bin5</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Bin6</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

L2 = 23 / 26
Constructing the L1/L2 SATT Frontier

![Bar chart and line graph showing frequency and number of observations dropped.]

- **Bin 1**: 5 observations (Control), 4 observations (Treatment)
- **Bin 2**: 6 observations (Control), 7 observations (Control)
- **Bin 3**: 2 observations (Control), 3 observations (Treatment)
- **Bin 4**: 3 observations (Control), 4 observations (Treatment)
- **Bin 5**: 3 observations (Control)
- **Bin 6**: 1 observation (Control), 2 observations (Treatment)

![Line graph showing the number of observations dropped against L2.]

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

**Legend**: Blue bars for Treatment, Red bars for Control
Constructing the L1/L2 SATT Frontier

![Graph showing frequency distribution and relationship with L2 and number of observations dropped.]

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

- **Relationship**:
  - L2 decreases as the number of observations dropped increases.

Note: The graph illustrates the distribution of observations across different bins, with separate bars for treatment and control groups. The L2 parameter decreases as the number of observations dropped increases, suggesting a trend in the data.
Constructing the L1/L2 SATT Frontier

![Bar chart and line graph showing frequency and treatment control across bins.]

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

![Graph showing number of observations dropped over L2 values.]

- **Number of Observations Dropped** vs **L2 Value**
Constructing the L1/L2 SATT Frontier
Constructing the L1/L2 SATT Frontier

The diagram shows the frequency distribution across different bins for treatment and control groups.

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

The graph on the right illustrates the number of observations dropped against L2 values, indicating a decreasing trend as the number of observations increases.
Constructing the L1/L2 SATT Frontier

![Bar chart showing frequency and treatment control for each bin.

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

A graph showing the number of observations dropped vs. L2.

- Number of observations dropped: 0 to 10
- L2: 0.00 to 0.12

Legend:
- Blue square: Treatment
- Red square: Control

L2 / Number of Observations Dropped
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
Conclusions

- The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier

- 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
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 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
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