Quantitative Social Science Methods, I, Lecture Notes: Detecting and Reducing Model Dependence in Causal Inference

Gary King[1]
Institute for Quantitative Social Science
Harvard University

August 17, 2020

[1] GaryKing.org

Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

The Matching Frontier

Readings in Model Dependence

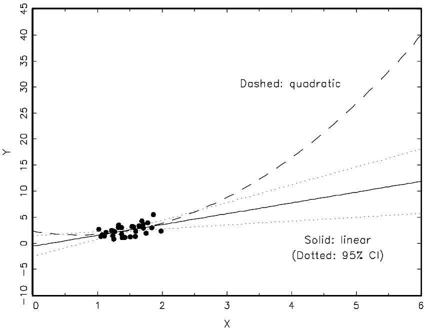
- King, Gary and Langche Zeng. "The Dangers of Extreme Counterfactuals," *Political Analysis*, 14, 2, (2007): 131-159.
- King, Gary and Langche Zeng. "When Can History be Our Guide? The Pitfalls of Counterfactual Inference," *International Studies Quarterly*, 2006, 51 (March, 2007): 183–210.
- Related Software: WhatIf, MatchIt, Zelig, CEM

`j.mp/causalinference`

Counterfactuals

- Three types:
  1. Forecasts What will the mortality rate be in 2025?
  2. Whatif Questions What would have happened if the U.S. had not invaded Iraq?
  3. Causal Effects What is the causal effect of the Iraq war on World GDP? (a factual minus a counterfactual)
- Counterfactuals are part of most social science research

Which model would you choose? (Both fit the data well.)



- Compare prediction at $x = 1.5$ to prediction at $x = 5$
- How do you choose a model? R^2? Some "test"? "Theory"?
- The bottom line: answers to some questions don't exist in the data. We show how to determine which ones.
- Same for what if questions, predictions, and causal inferences

Model Dependence Proof

Model Free Inference

To estimate $E(Y|X = x)$ at x, average many observed Y with value x

Assumptions (Model-Based Inference)

1. Definition: model dependence at x is the difference between predicted outcomes for any two models that fit about equally well.
2. The functional form follows strong continuity (think smoothness, although it is less restrictive)

Result

The maximum degree of model dependence: a function of the distance from the counterfactual to the data

A Simple Measure of Distance from The Data

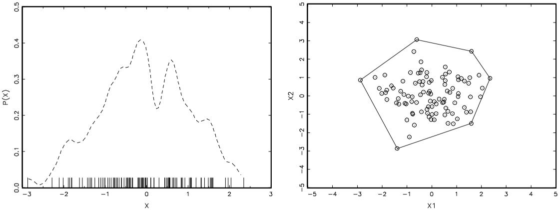


Figure: The Convex Hull

- Interpolation: Inside the convex hull
- Extrapolation: Outside the convex hull
- Works mathematically for any number of X variables
- Software to determine whether a point is in the hull (which is all we need) without calculating the hull (which would take forever), so its fast; see `GaryKing.org/whatif`

Model Dependence Example

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

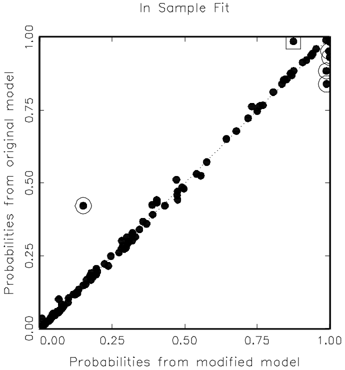
- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,...
- Counterfactual question: Switch UN intervention for each war
- Data analysis: Logit model
- The question: How *model dependent* are the results?
- Percent of counterfactuals in the convex hull: 0%
- $\rightsquigarrow$ without estimating any models, we know: inferences will be model dependent
- For illustration: let's find an example....

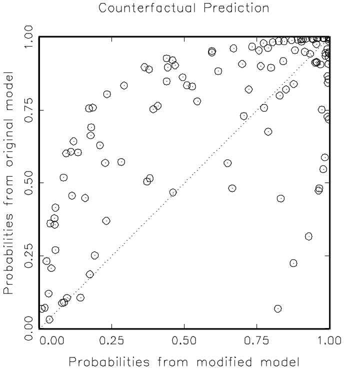
Two Logit Models, Apparently Similar Results

Effect of Multilateral UN Intervention on Peacebuilding Success

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

Model Dependence: Same Fit, Different Predictions





Detecting Model Dependence

Matching to Reduce Model Dependence

Three Matching Methods

Problems with Propensity Score Matching

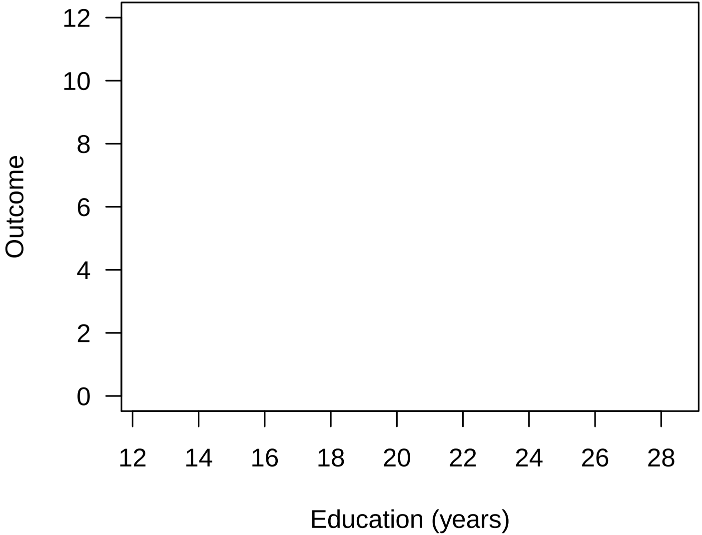
The Matching Frontier

Readings, Matching

- Do powerful methods have to be complicated?
  - “Causal Inference Without Balance Checking: Coarsened Exact Matching” (*PA*, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)
- The most popular method (propensity score matching, used in 140,000 articles!) sounds magical:
  - “Why Propensity Scores Should Not Be Used for Matching” (Gary King, Richard Nielsen) (*PA*, 2019; Gary King and Richard Nielsen)
- Matching methods optimize either imbalance ($\approx$ bias) or # units pruned ($\approx$ variance); users need both simultaneously’:
  - “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (*AJPS*, 2017; Gary King, Christopher Lucas and Richard Nielsen)
- Current practice, matching as preprocessing: violates current statistical theory. So let’s change the theory:
  - “A Theory of Statistical Inference for Matching Methods in Causal Research” (Stefano Iacus, Gary King, Giuseppe Porro)

Matching to Reduce Model Dependence

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



The Problems Matching Solves

Without Matching:With~~out~~ Matching:

Imbalance~~Im~~balance ⇝ Model Dependence ⇝ ~~Model Dependence~~ ⇝ Researcher discretion ⇝ ~~Researcher discretion~~ ⇝ Bias ⇝ ~~Bias~~

A central project of statistics: Automating away human discretion

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[eff]
- conscientious effort doesn't avoid biases (Banaji 2013)[acc]
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)[exprt]
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the

What's Matching?

- Notation: Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

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

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control
- Quantities of Interest
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \underset{i \in \{T_i = 1\}}{\text{Mean}} (\text{TE}_i)$$

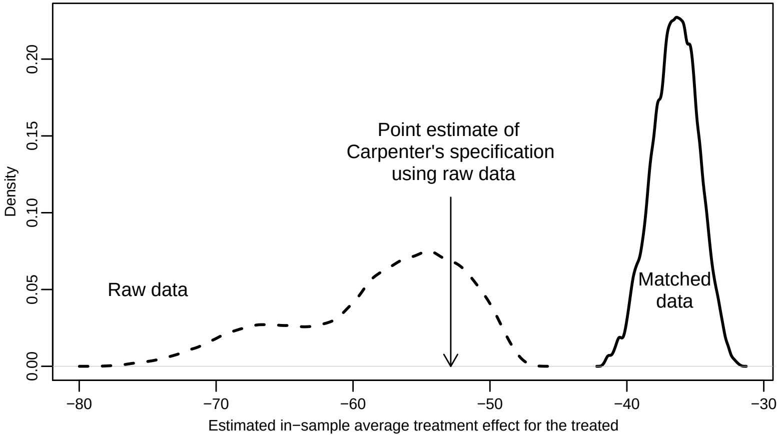
  2. FSATT: Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias

Evaluating Reduction in Model Dependence

Empirical Illustration: Carpenter, AJPS, 2002

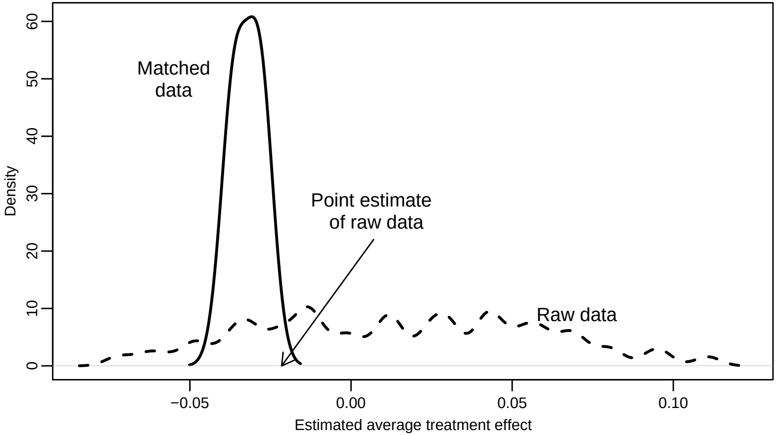
- Hypothesis: Democratic senate majorities slow FDA drug approval time
- Data: n = 408 new drugs (262 approved, 146 pending)
- Measured confounders: 18 (clinical factors, firm characteristics, media variables, etc.)
- Model: lognormal survival
- QOI: Causal effect of Democratic Senate majority (identified by Carpenter as not robust)
- Match: prune 49 units (2 treated, 17 control units)
- Run: 262,143 possible specifications; calculate SATT for each
- Evaluate: *Variability* in SATT across specifications
- (Normally we'd only use one or a few specifications)

Reducing Model Dependence



SATT Histogram: Effect of Democratic Senate majority on FDA drug approval time, across $262,143$ specifications

Another Example: Jeffrey Koch, AJPS, 2002



SATT Histogram: Effect of being a highly visible female Republican candidate across 63 possible specifications with the Koch data

Assumptions to Justify Current Practice

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

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

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

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

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
- Other matching methods dominate PSM (wait, it gets worse)

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

Procedure

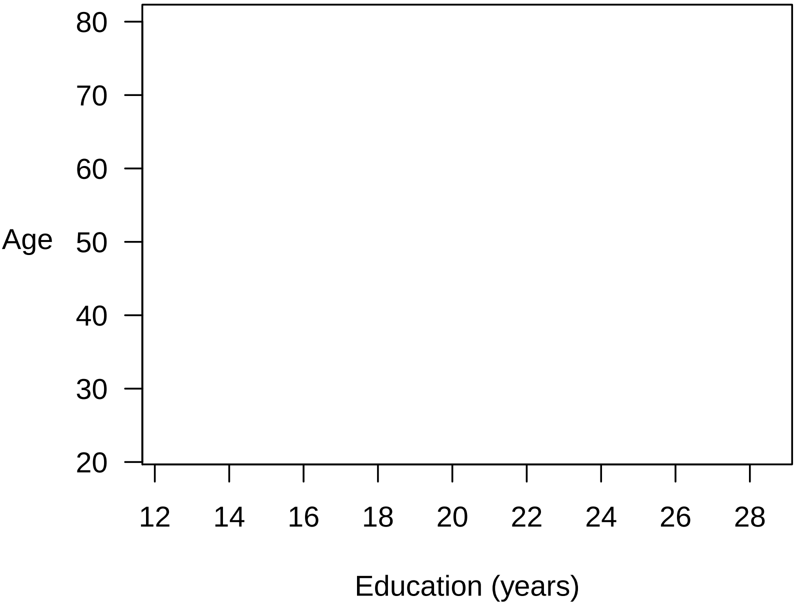
1. Preprocess (Matching)
    - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
    - Match each treated unit to the nearest control unit
    - Control units: not reused; pruned if unused
    - Prune matches if Distance>*caliper*
    - (Many adjustments available to this basic method)
2. Estimation Difference in means or a model

Interpretation

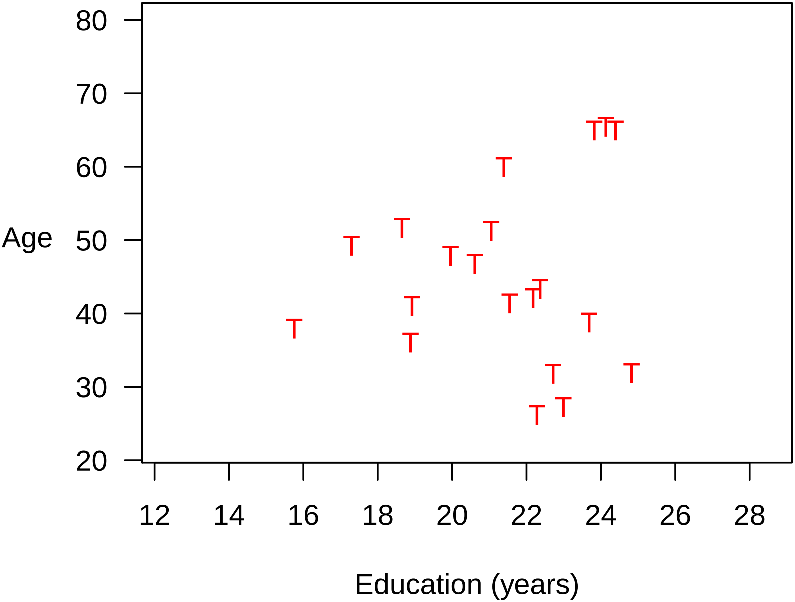
- Quiz: Do you understand the distance trade offs?
- Quiz: Does standardization help?

$\rightsquigarrow$ Mahalanobis is for methodologists; in applications, use Euclidean!

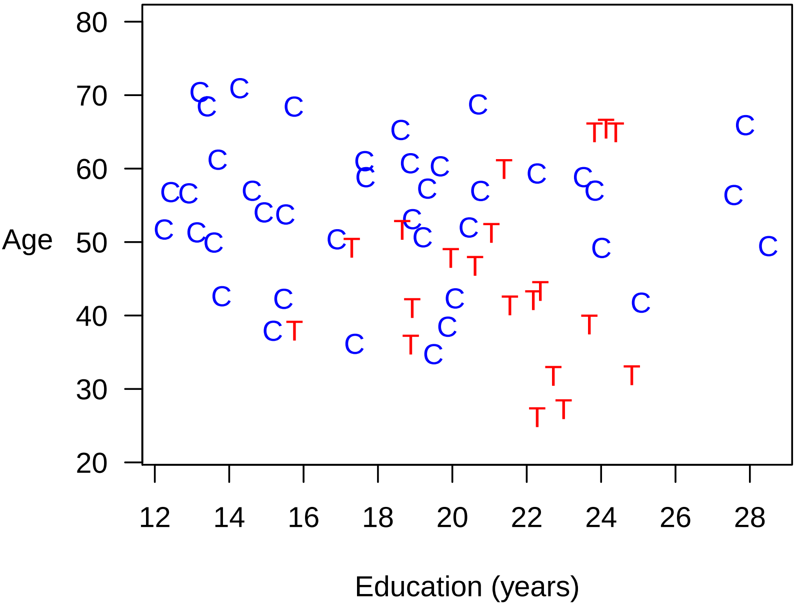
Mahalanobis Distance Matching



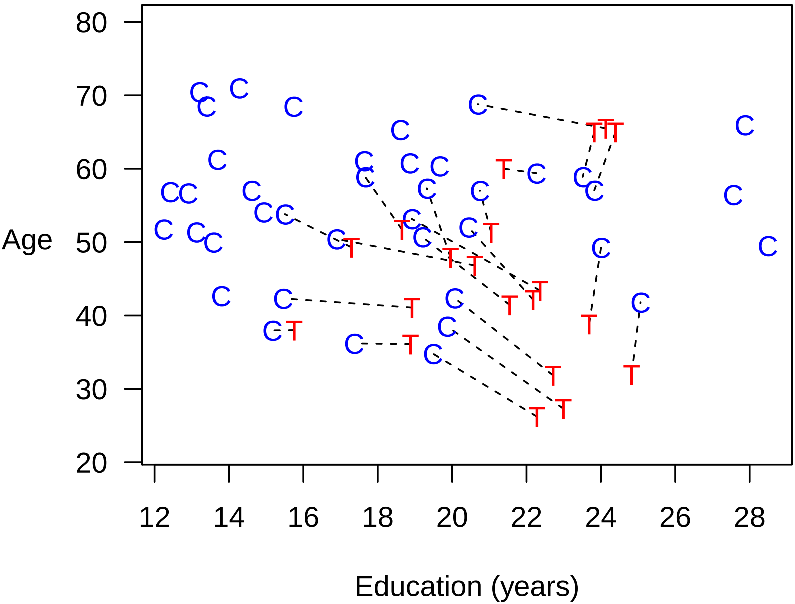
Mahalanobis Distance Matching



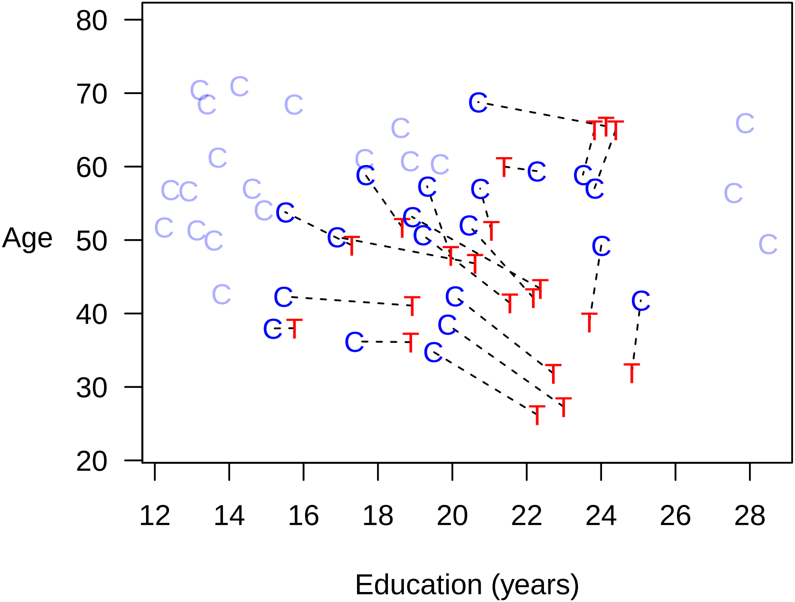
Mahalanobis Distance Matching



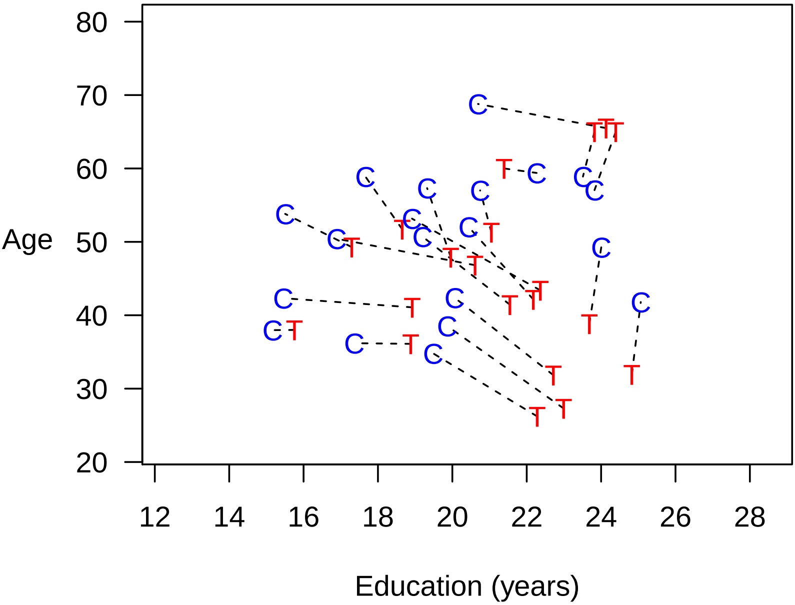
Mahalanobis Distance Matching



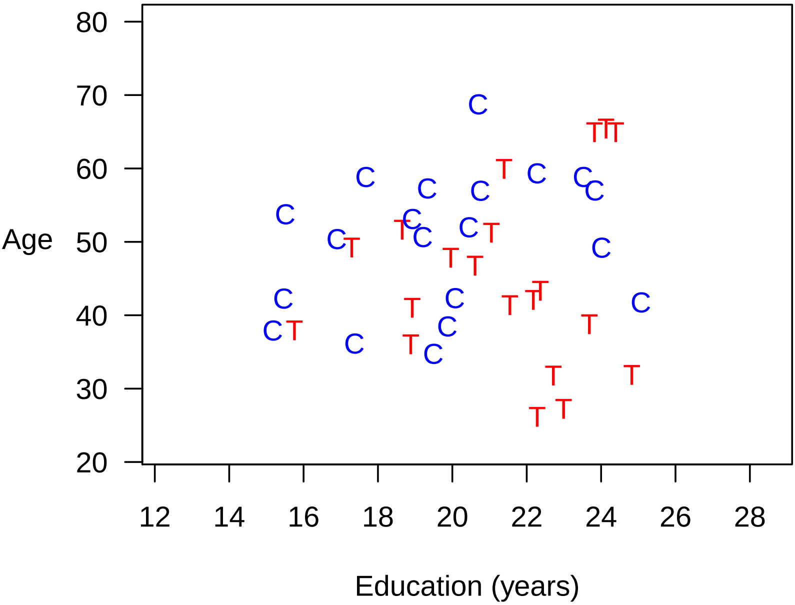
Mahalanobis Distance Matching



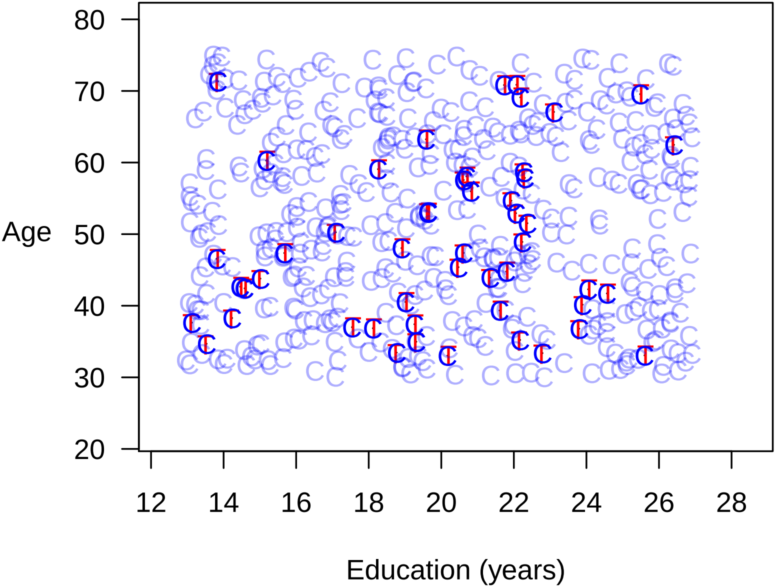
Mahalanobis Distance Matching



Mahalanobis Distance Matching



Best Case: Mahalanobis Distance Matching



Method 2: Coarsened Exact Matching

(Approximates Fully Blocked Experiment)

Procedure

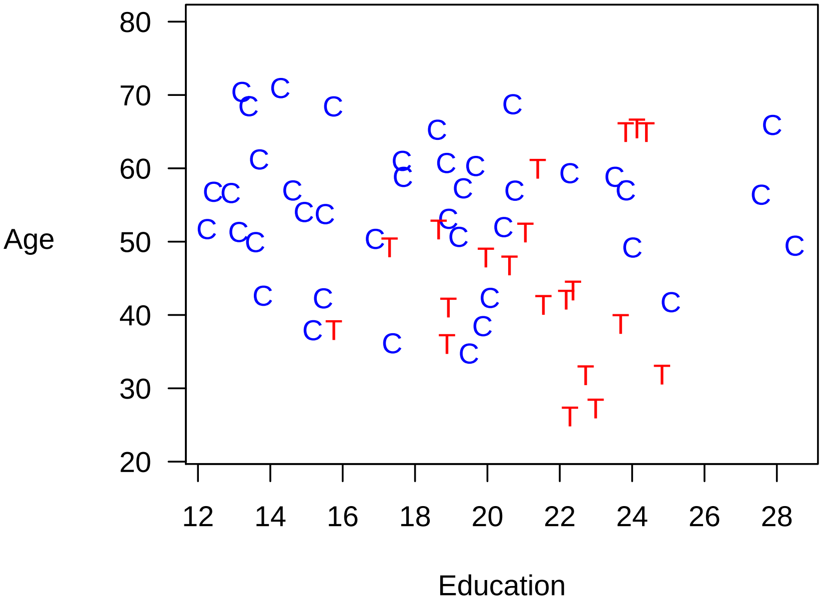
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with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

Interpretation

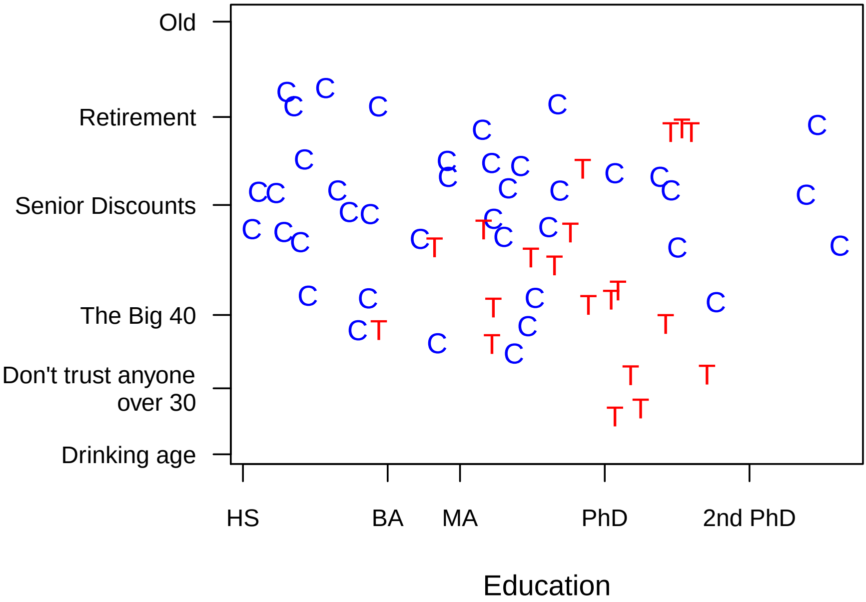
- Quiz: Do you understand distance trade offs?
- Quiz: What do you do if you have too few observations?

Coarsened Exact Matching

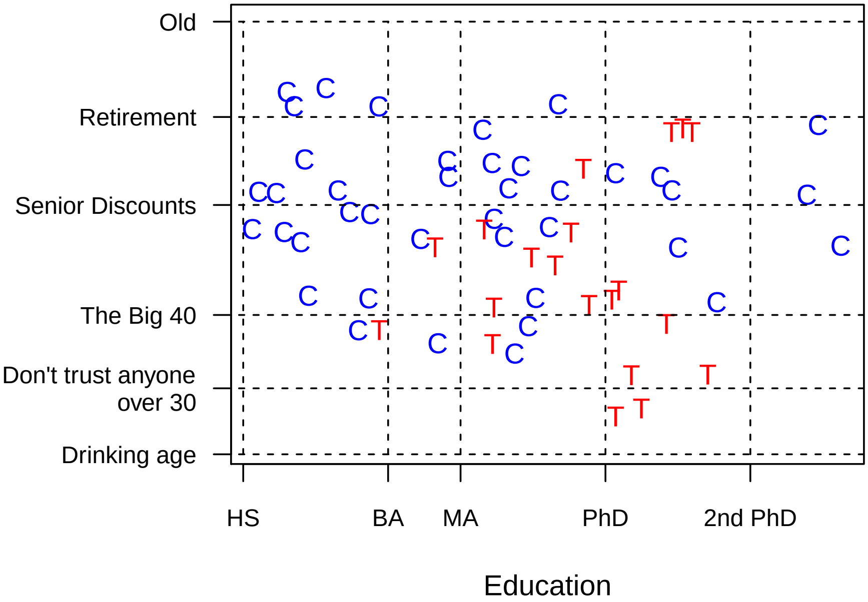
Coarsened Exact Matching



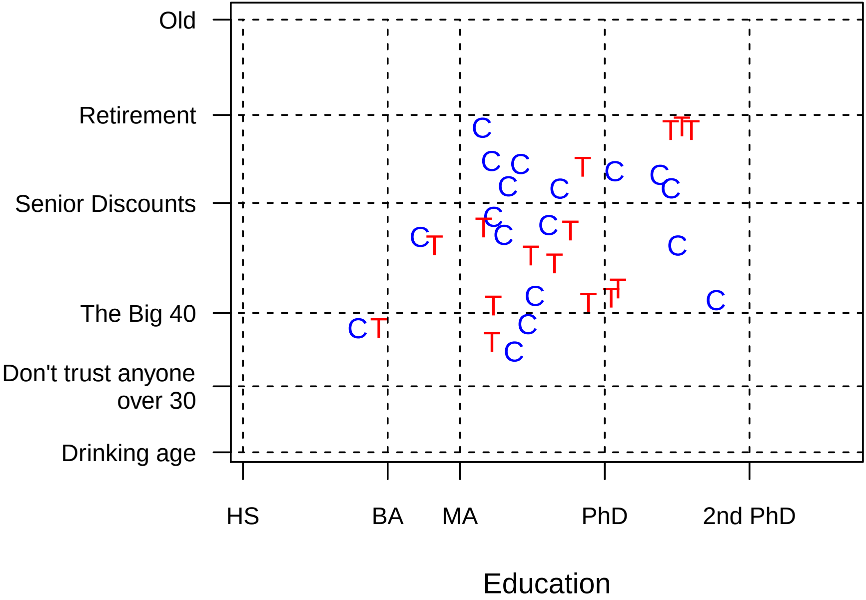
Coarsened Exact Matching



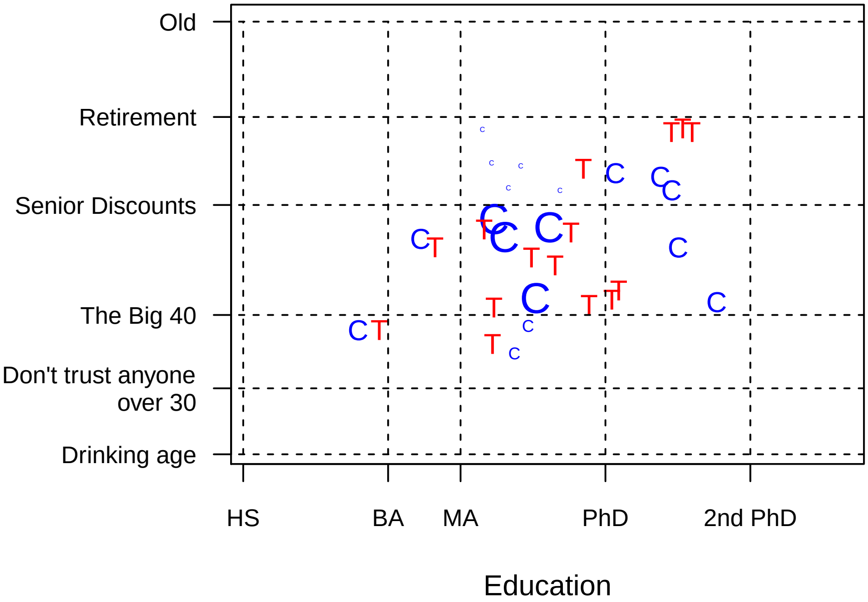
Coarsened Exact Matching



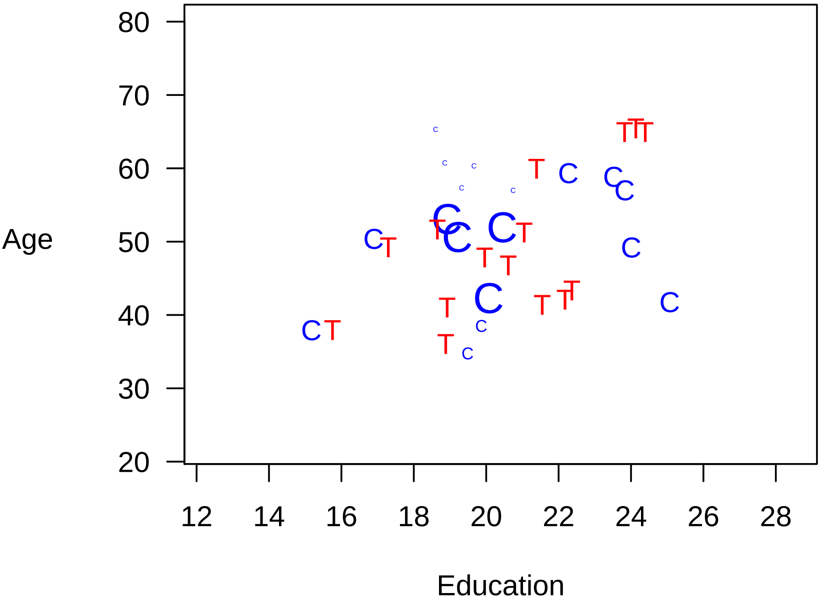
Coarsened Exact Matching



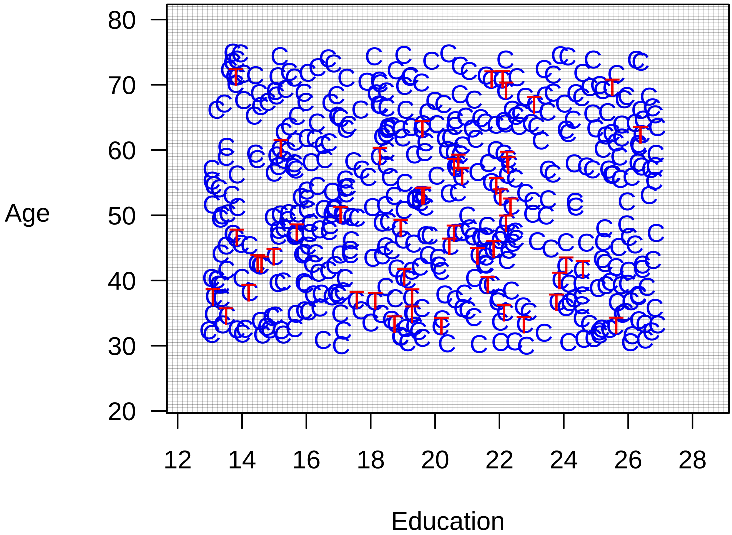
Coarsened Exact Matching

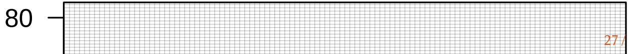


Coarsened Exact Matching



Best Case: Coarsened Exact Matching





Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

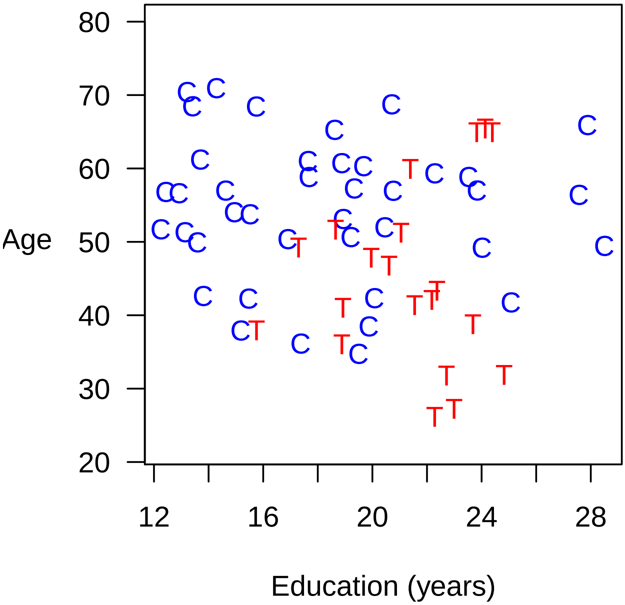
Procedure

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance>*caliper*
   - (Many adjustments available to this basic method)
2. Estimation Difference in means or a model

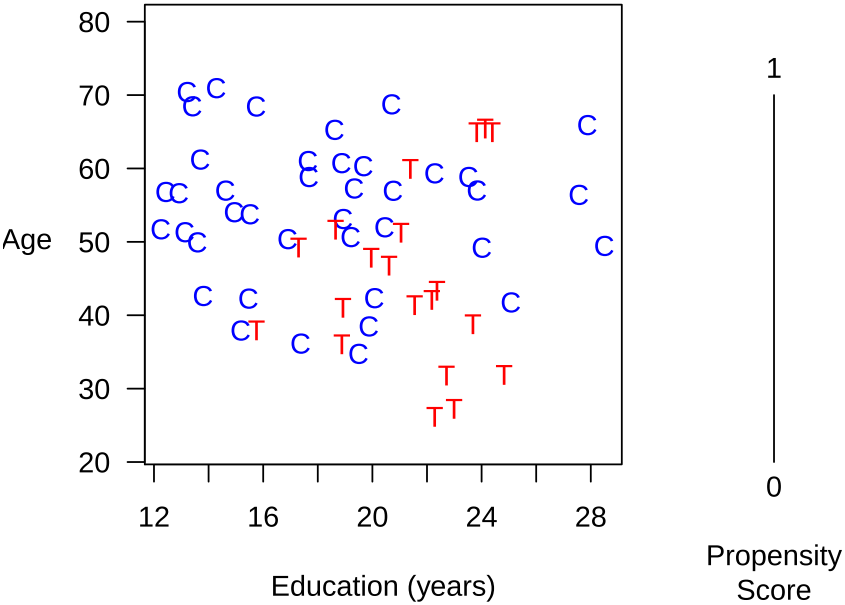
Interpretation

- Quiz: Do you understand distance trade offs?
- Quiz: What do you do when one variable is very important?

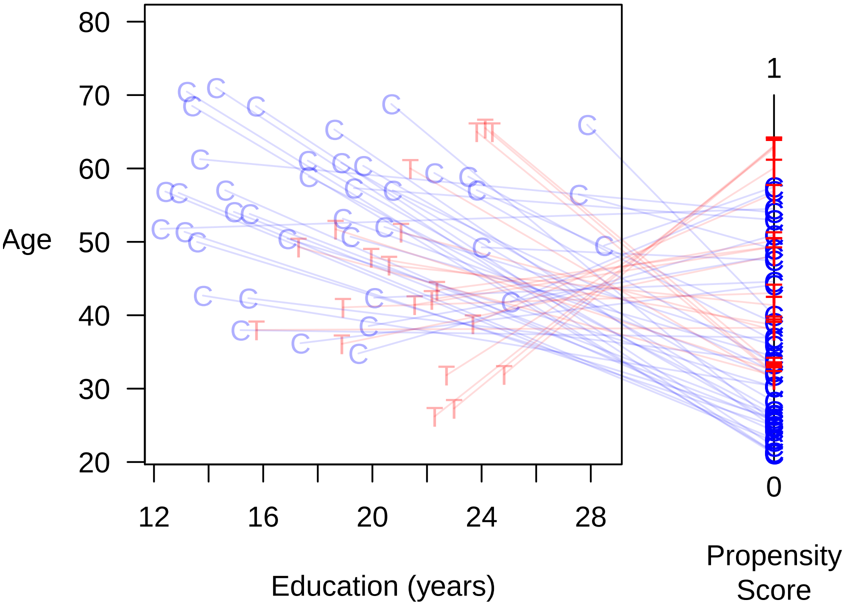
Propensity Score Matching



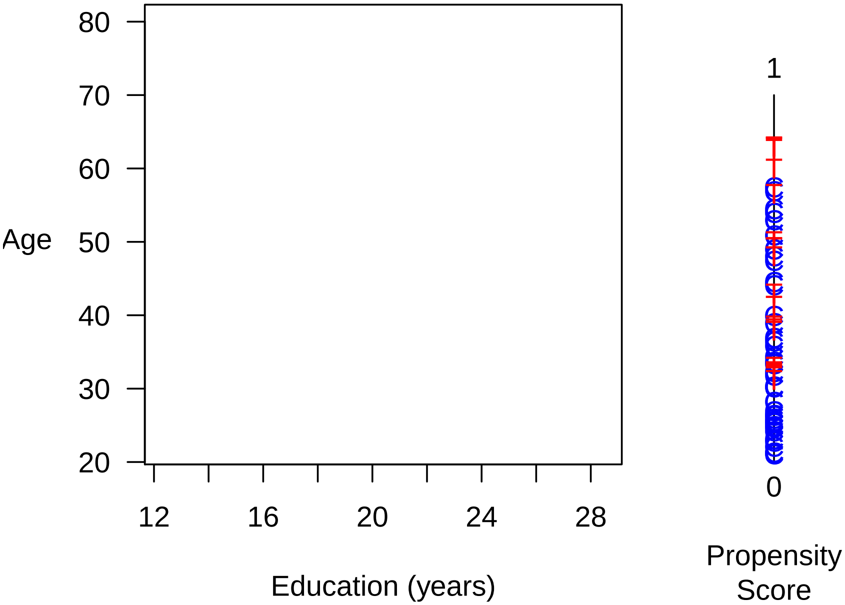
Propensity Score Matching



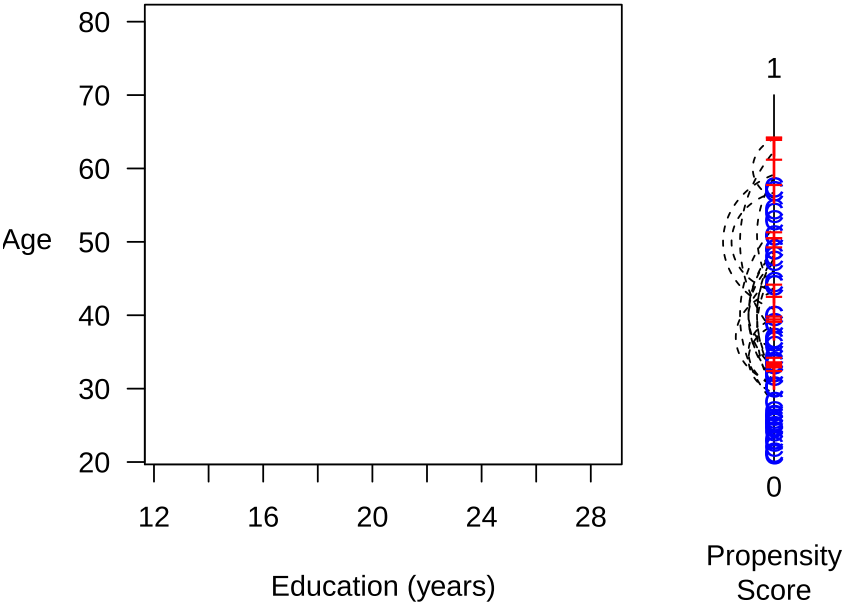
Propensity Score Matching



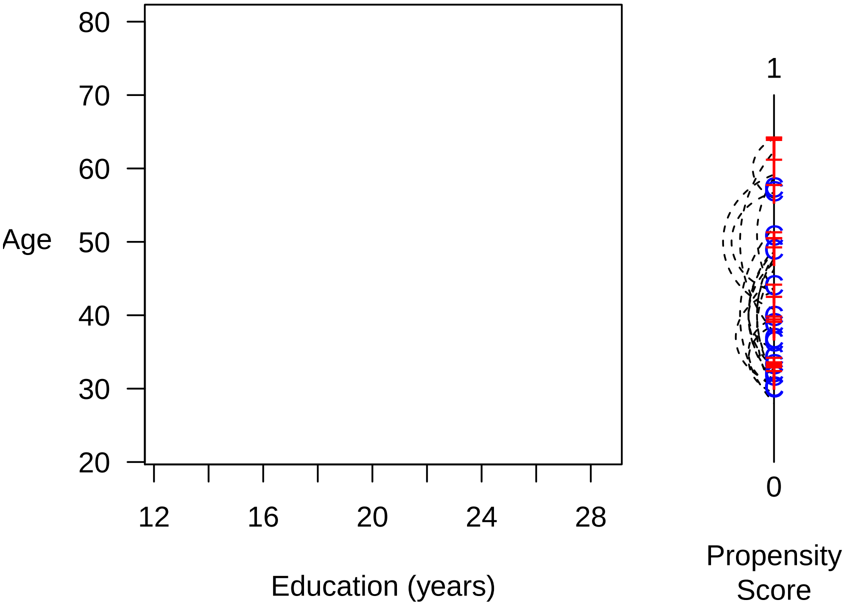
Propensity Score Matching



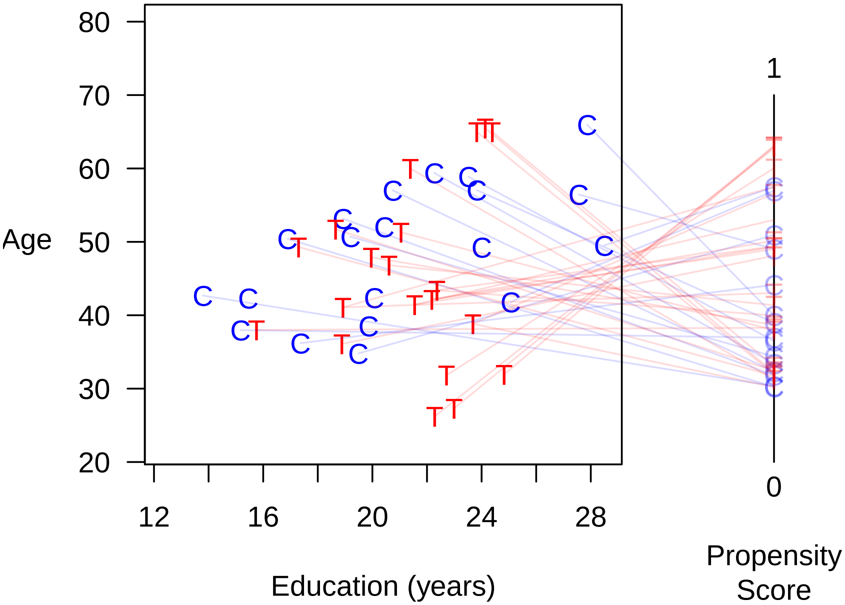
Propensity Score Matching



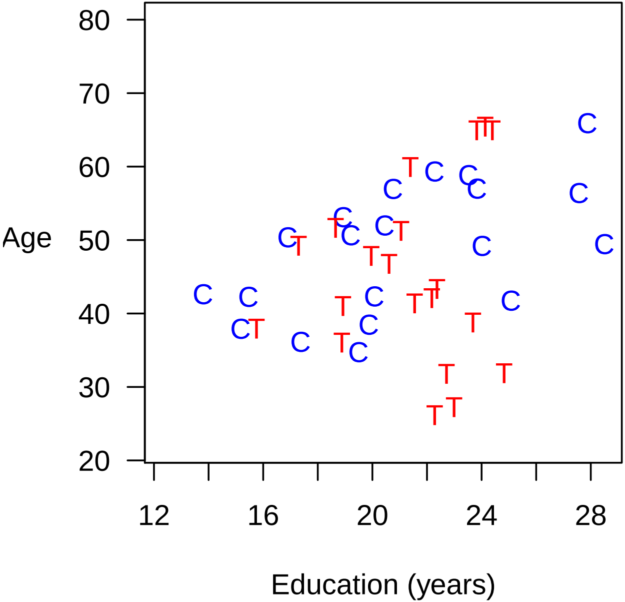
Propensity Score Matching



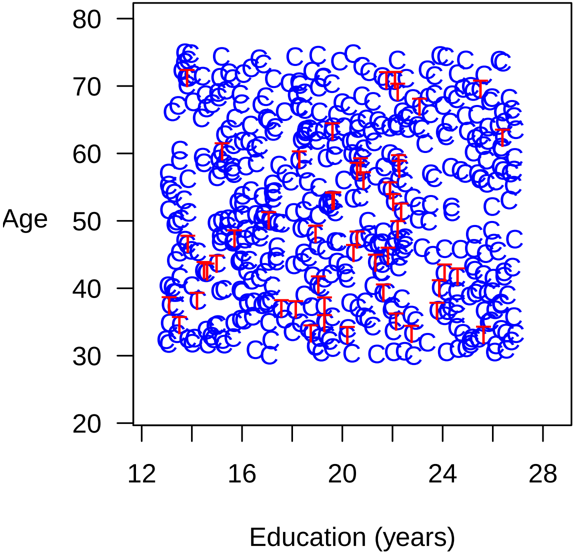
Propensity Score Matching

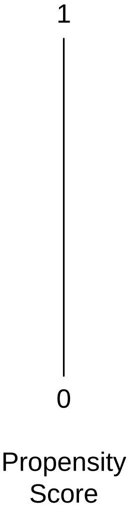


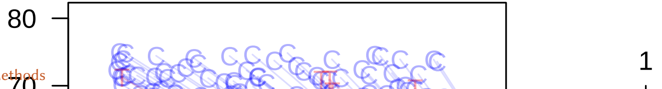
Propensity Score Matching



Best Case: Propensity Score Matching is Suboptimal







Random Pruning Increases Imbalance

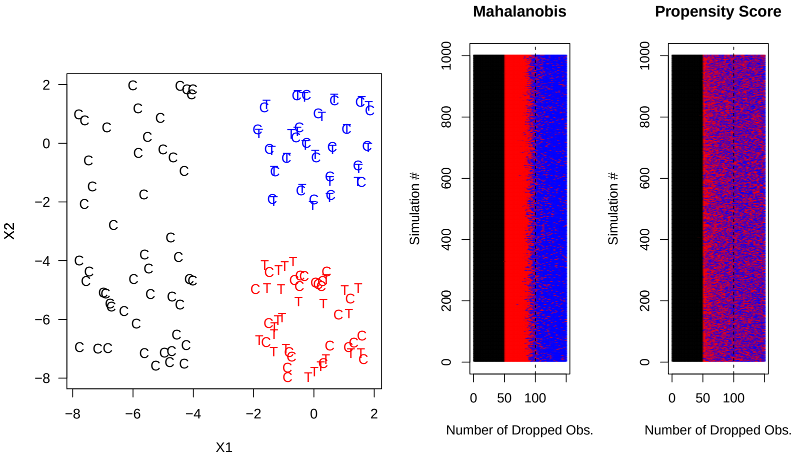
Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets: 2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow$ n declines $\rightsquigarrow$ $E(d^2)$ increases
  - $\Longrightarrow$ random pruning increases imbalance
- Result is completely general

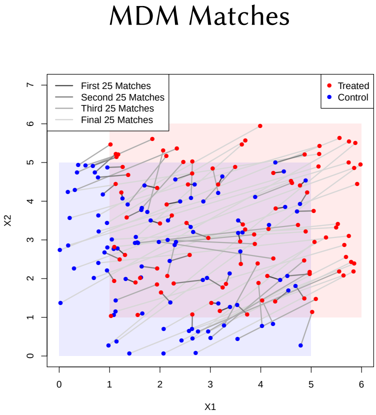
PSM's Statistical Properties

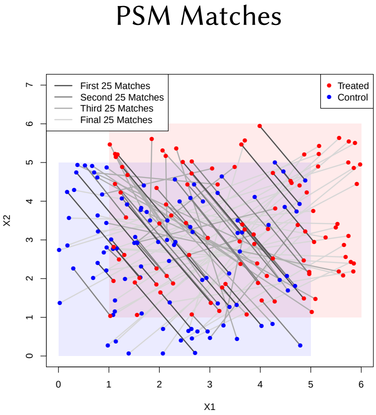
1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     $X_c = X_t \Longrightarrow \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Rightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
   - If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
   - Doesn't PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates

PSM is Blind Where Other Methods Can See



What Does PSM Match?

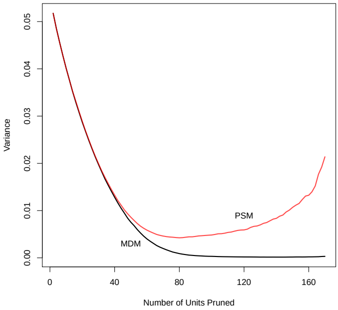




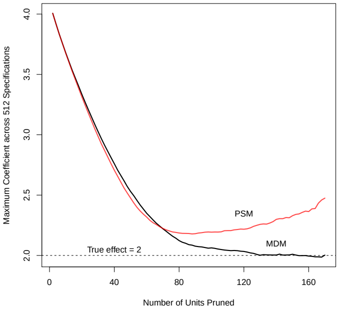
Controls: $X_1, X_2 \sim \text{Uniform}(0,5)$
Treateds: $X_1, X_2 \sim \text{Uniform}(1,6)$

PSM Increases Model Dependence & Bias

Model Dependence



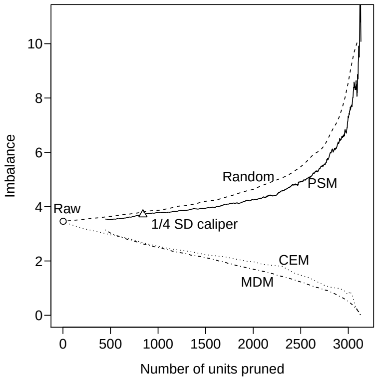
Bias



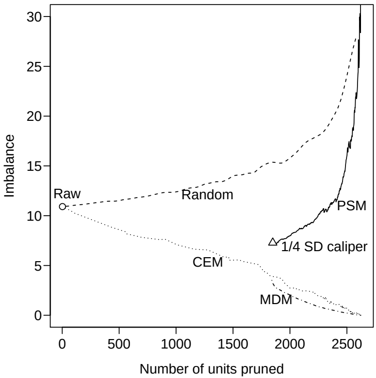
$$Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i$$
$$\epsilon_i \sim N(0, 1)$$

The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012)



Nielsen et al. (AJPS, 2011)

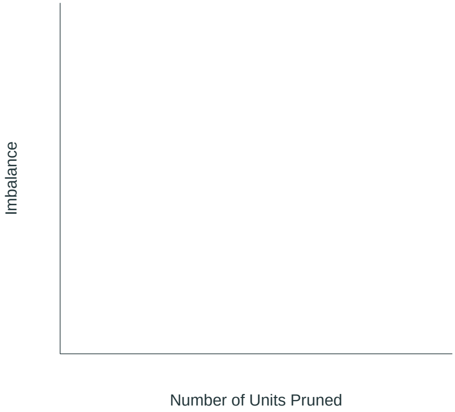


Similar pattern for > 20 other real data sets we checked

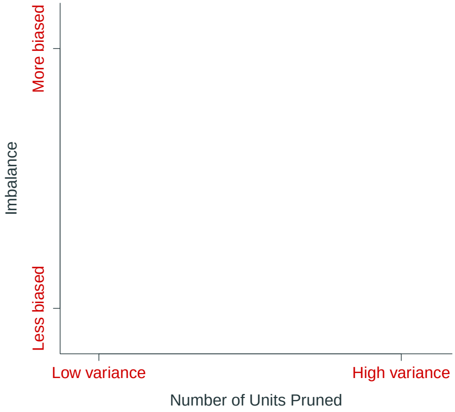
Tensions in Existing Matching Methods

- Maximize one metric; judge against another: Propensity score matching, compared with var-by-var diff in means
- Choose n; check imbalance after: Propensity score matching, Mahalanobis
- Choose imbalance; check n after: exact matching, CEM

A Solution: The Matching Frontier



A Solution: The Matching Frontier



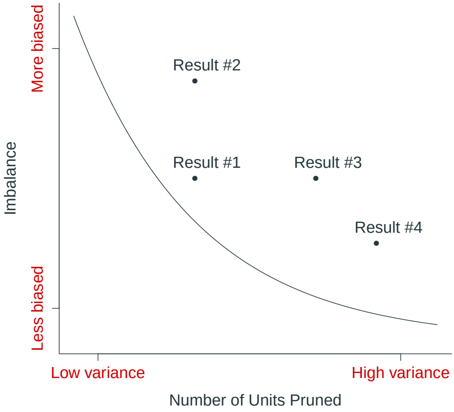
A Solution: The Matching Frontier

A Solution: The Matching Frontier

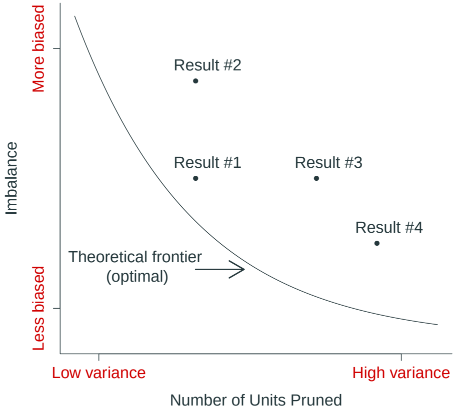
A Solution: The Matching Frontier

A Solution: The Matching Frontier

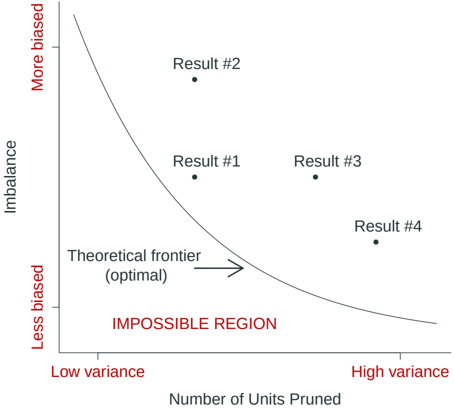
A Solution: The Matching Frontier



A Solution: The Matching Frontier



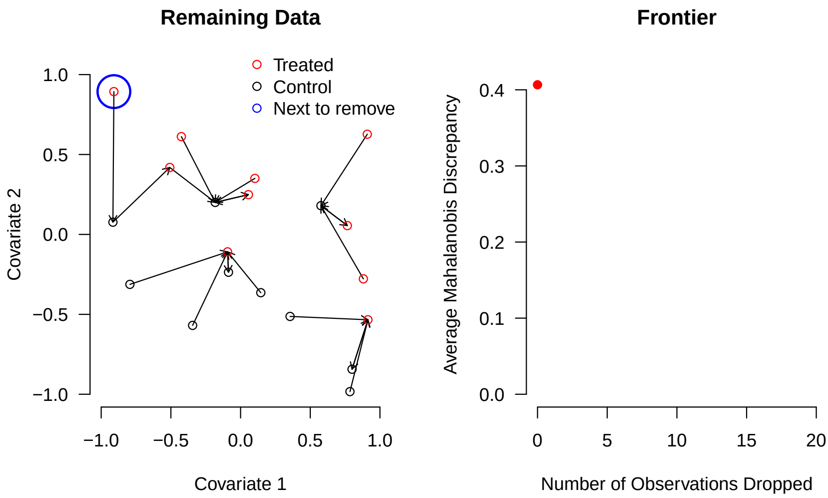
A Solution: The Matching Frontier

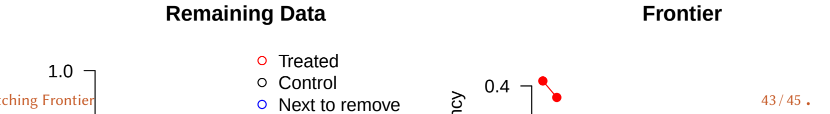


How hard is the frontier to calculate?

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

Constructing the FSATT Mahalanobis Frontier

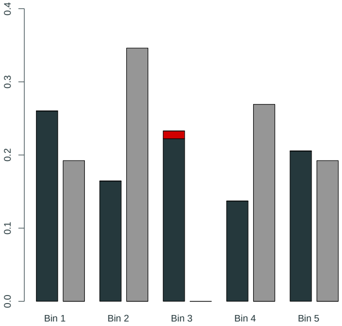




Discrete algorithm

Short version:

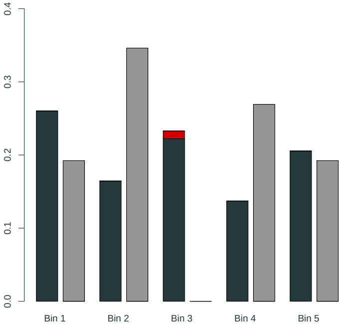
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

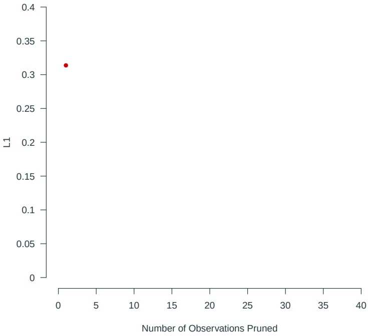


Discrete algorithm

Short version:

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

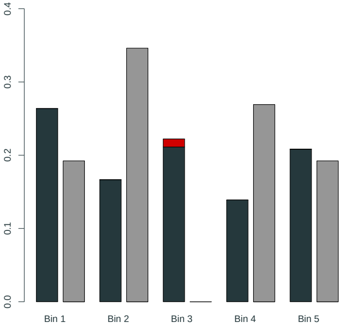


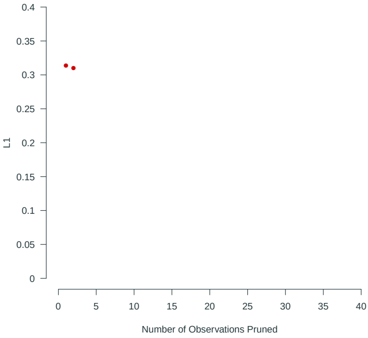


Discrete algorithm

Short version:

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

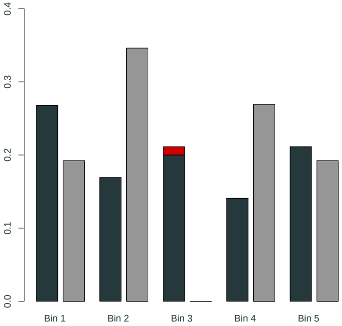


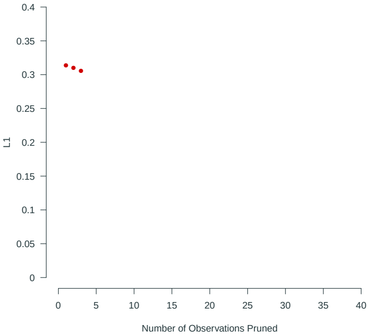


Discrete algorithm

Short version:

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

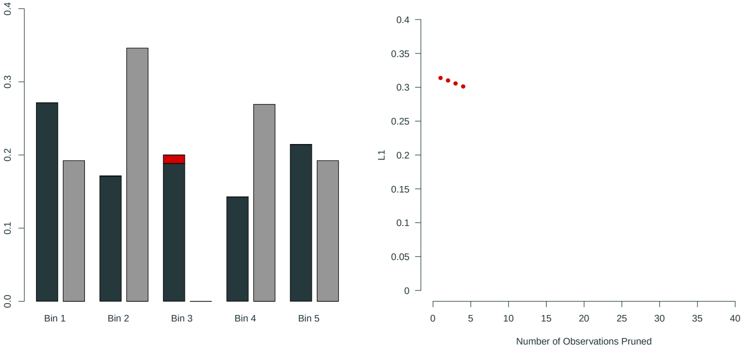




Discrete algorithm

Short version:

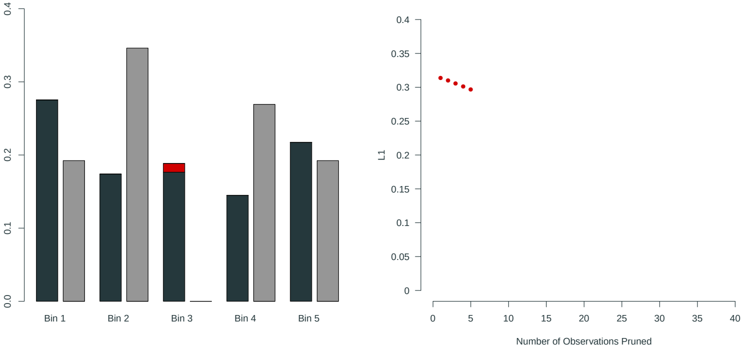
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units



Discrete algorithm

Short version:

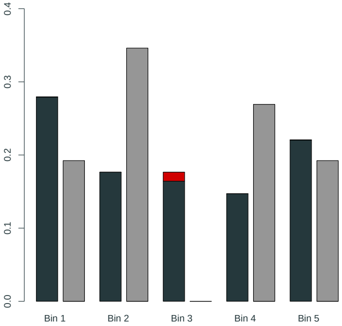
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

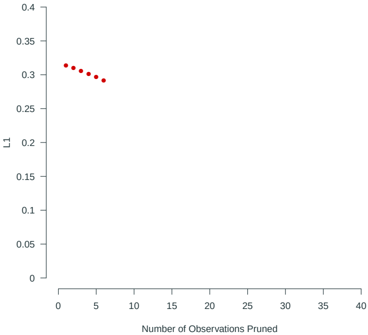


Discrete algorithm

Short version:

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

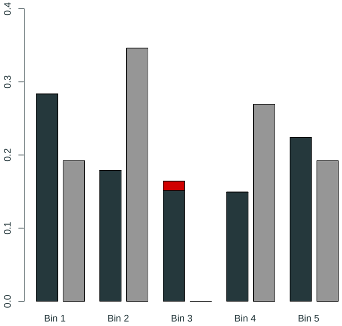


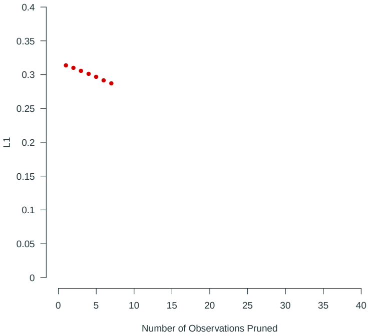


Discrete algorithm

Short version:

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

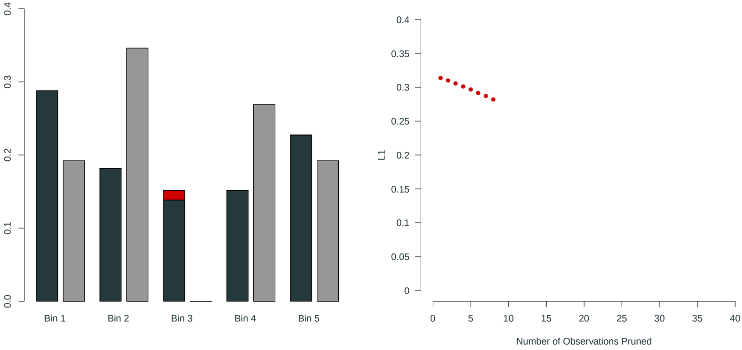




Discrete algorithm

Short version:

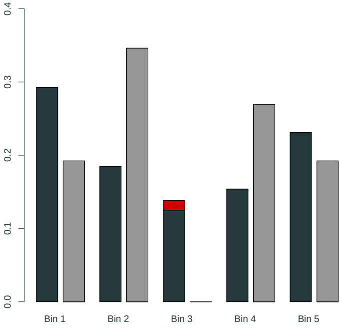
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

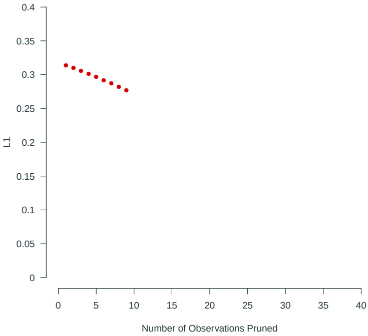


Discrete algorithm

Short version:

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

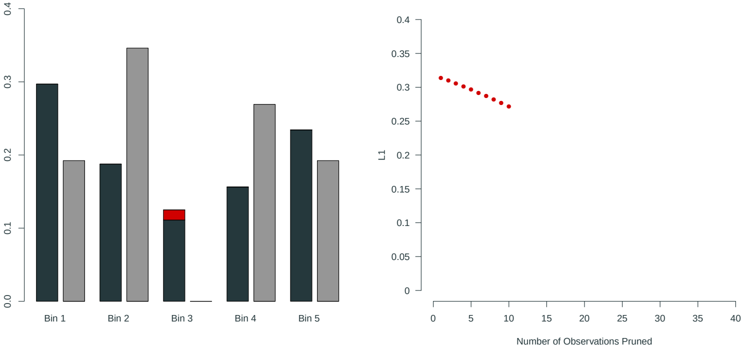




Discrete algorithm

Short version:

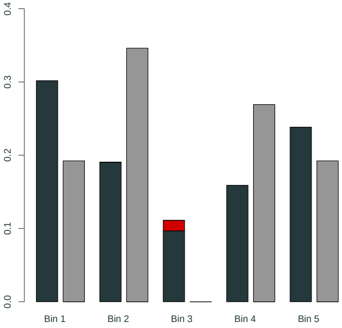
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

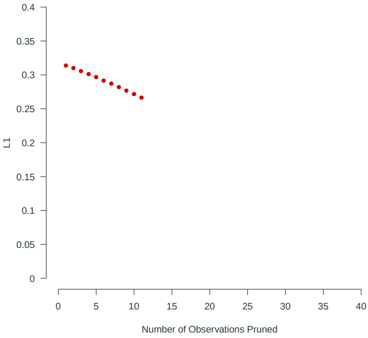


Discrete algorithm

Short version:

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

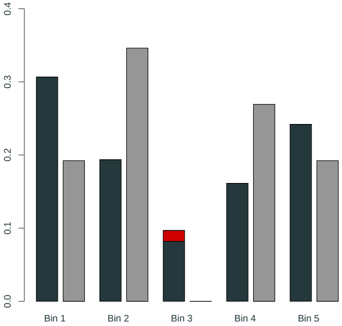


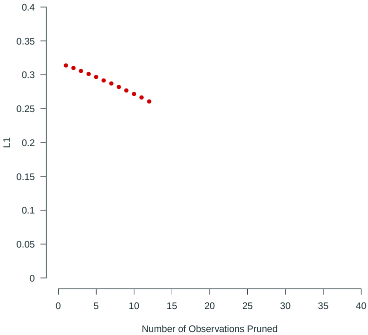


Discrete algorithm

Short version:

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

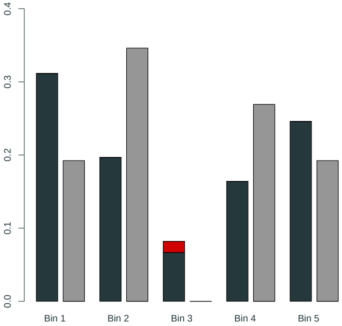


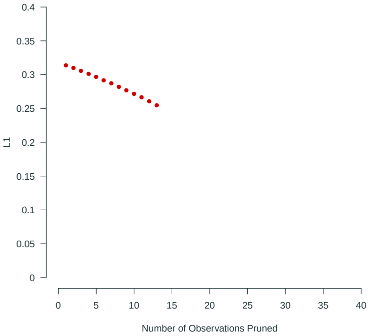


Discrete algorithm

Short version:

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

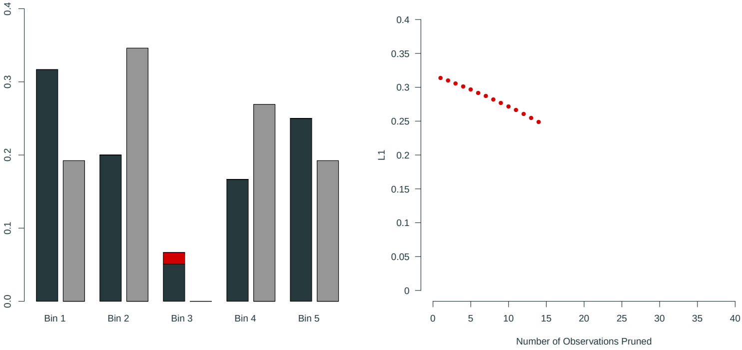




Discrete algorithm

Short version:

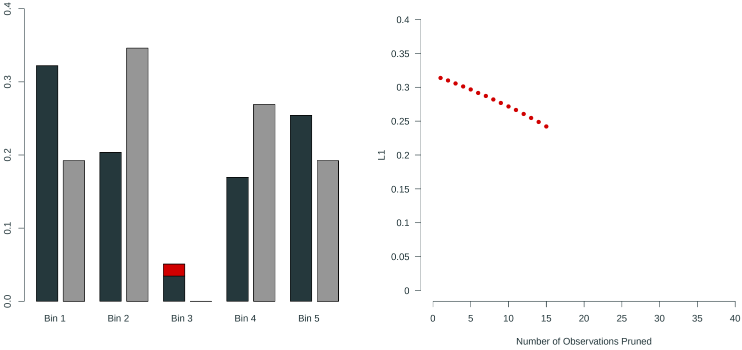
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units



Discrete algorithm

Short version:

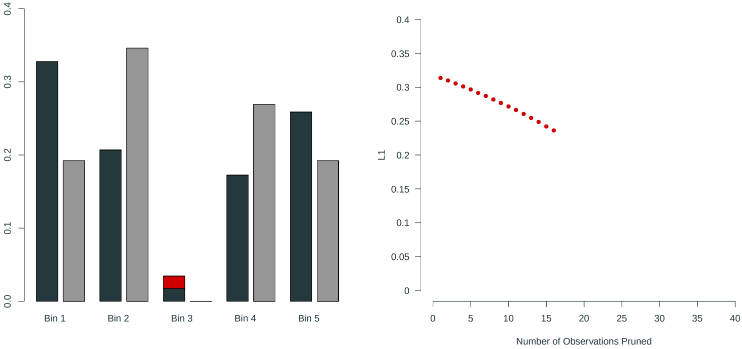
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units



Discrete algorithm

Short version:

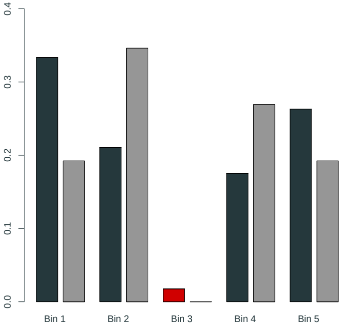
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

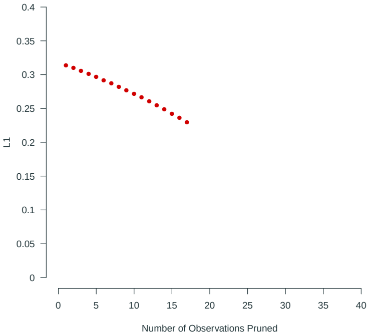


Discrete algorithm

Short version:

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

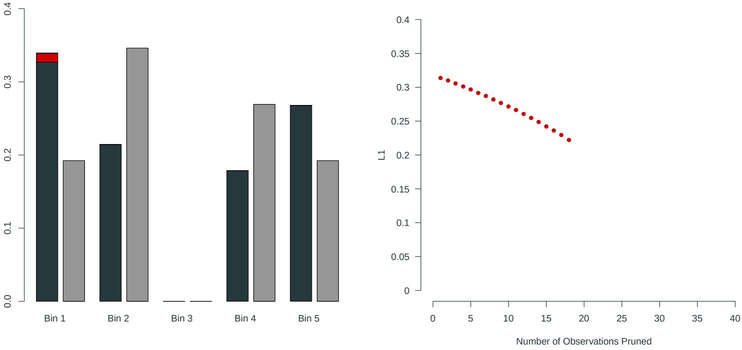




Discrete algorithm

Short version:

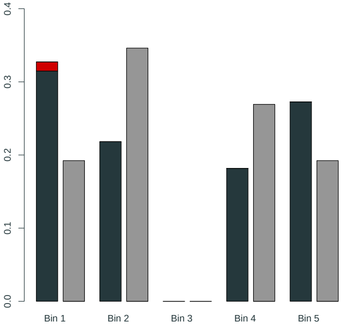
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

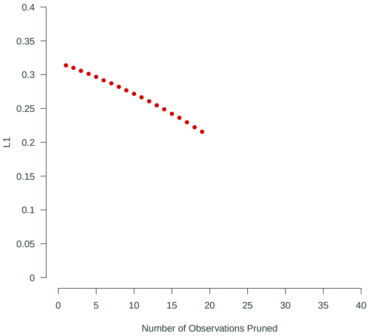


Discrete algorithm

Short version:

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

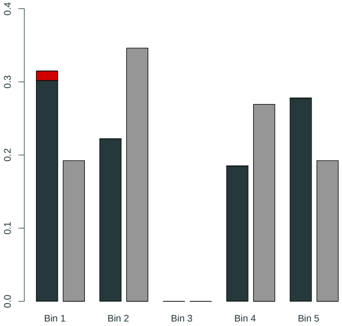


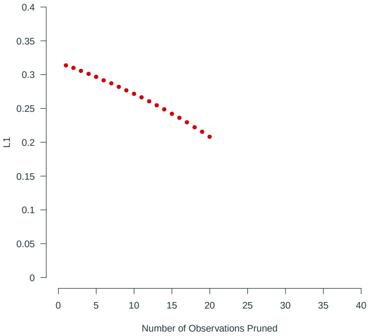


Discrete algorithm

Short version:

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

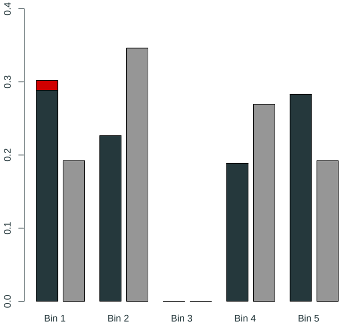


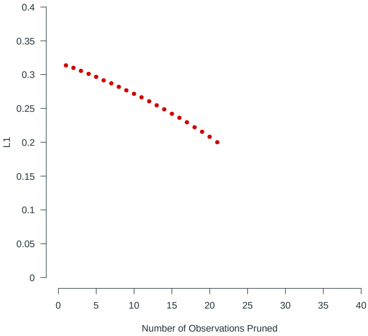


Discrete algorithm

Short version:

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

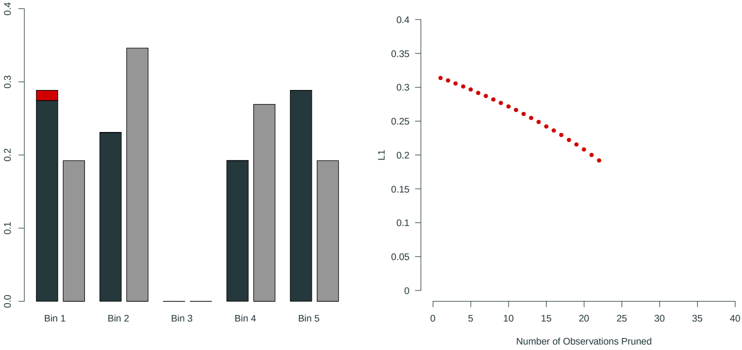




Discrete algorithm

Short version:

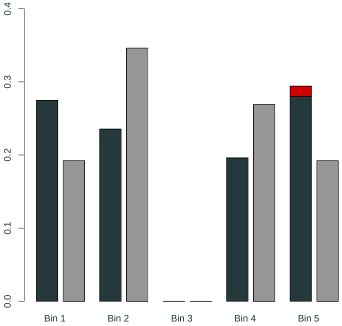
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

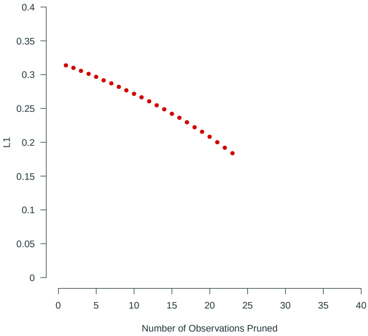


Discrete algorithm

Short version:

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

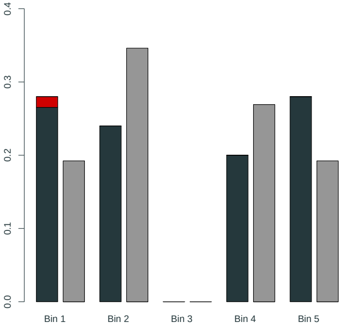


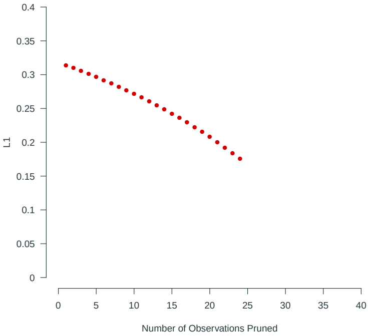


Discrete algorithm

Short version:

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

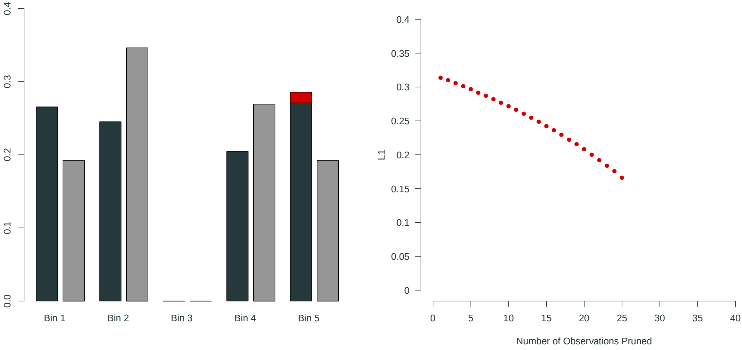




Discrete algorithm

Short version:

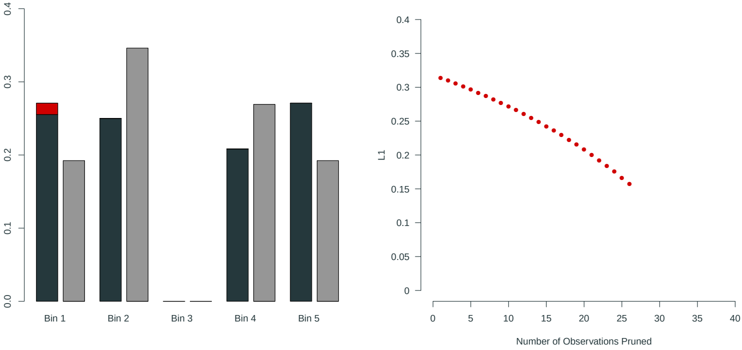
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units



Discrete algorithm

Short version:

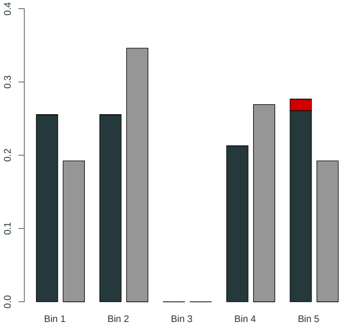
- Calculate bins
- Until balance stops improving, greedily prune a control unit from the bin with the largest proportional difference between control and treated units

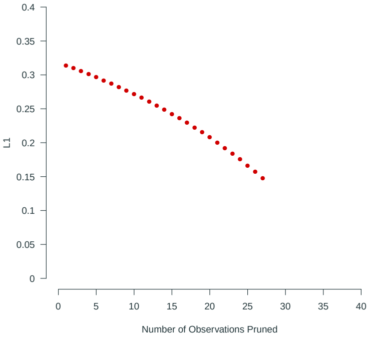


Discrete algorithm

Short version:

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

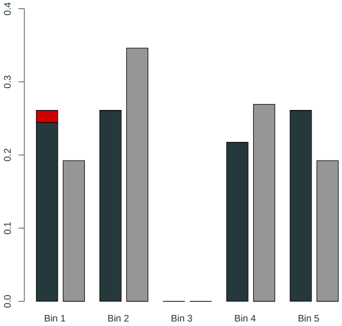


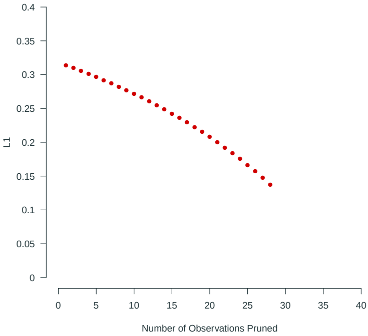


Discrete algorithm

Short version:

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

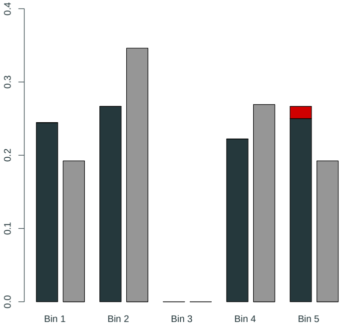


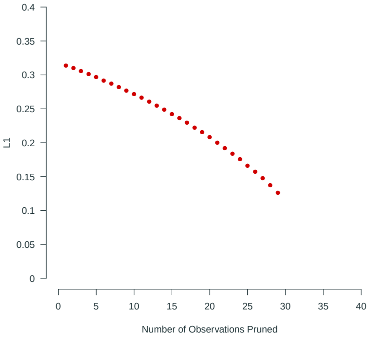


Discrete algorithm

Short version:

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

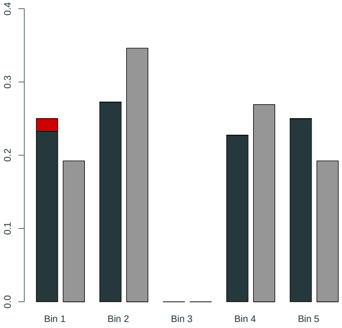


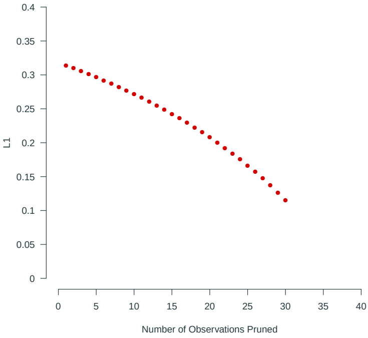


Discrete algorithm

Short version:

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

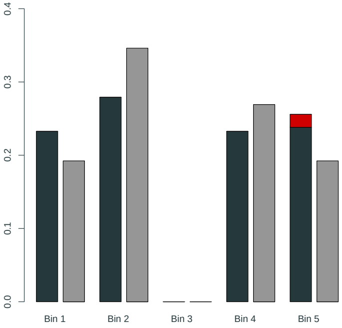


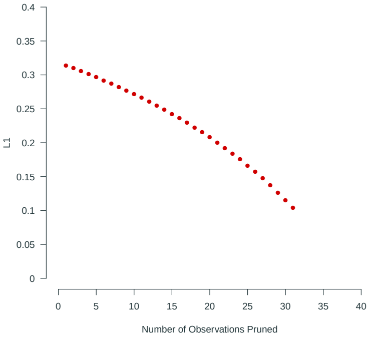


Discrete algorithm

Short version:

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

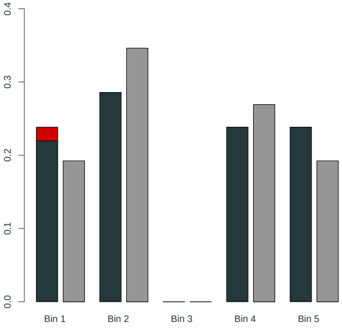


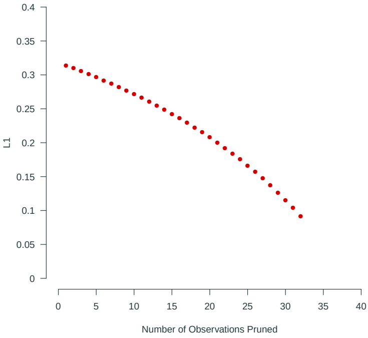


Discrete algorithm

Short version:

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

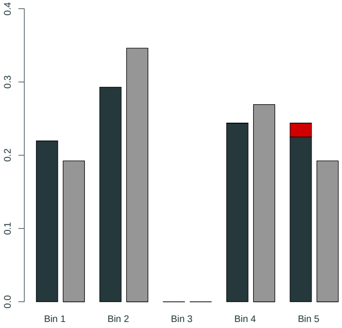


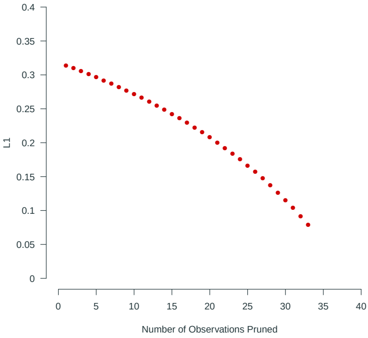


Discrete algorithm

Short version:

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

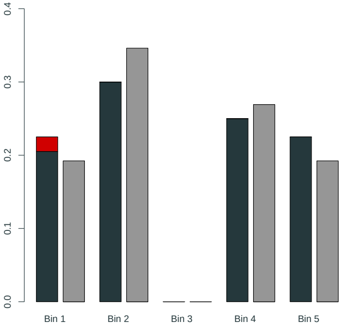


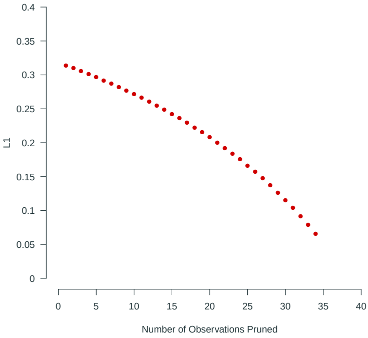


Discrete algorithm

Short version:

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

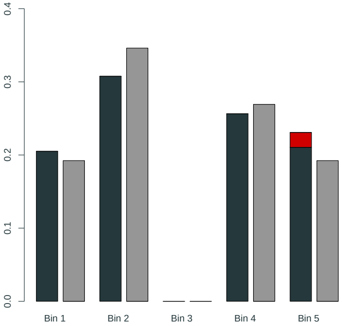


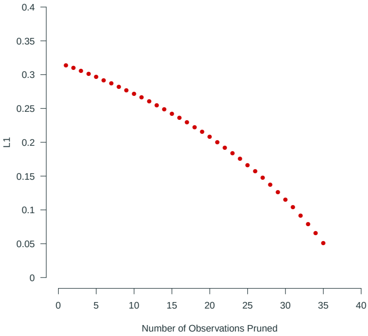


Discrete algorithm

Short version:

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

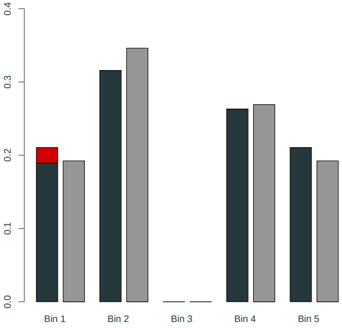


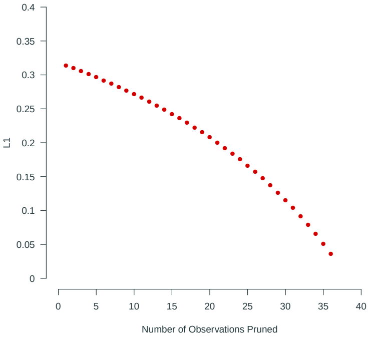


Discrete algorithm

Short version:

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

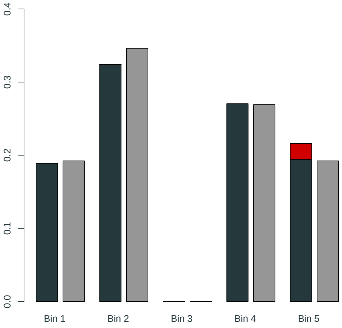


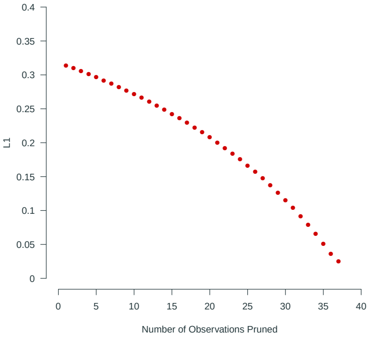


Discrete algorithm

Short version:

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

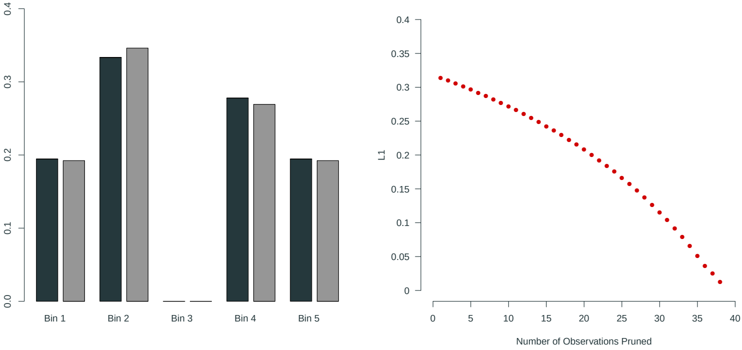




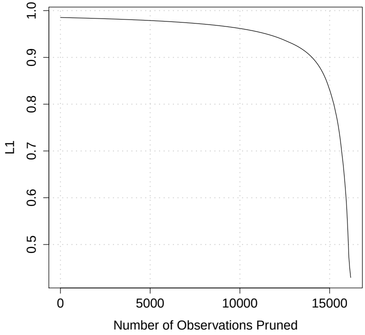
Discrete algorithm

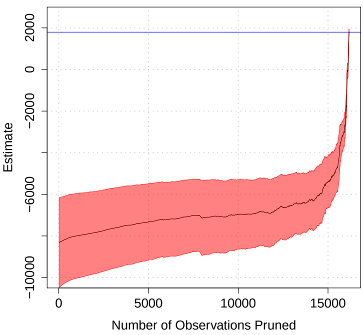
Short version:

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



Job Training Data: Frontier and Causal Estimates





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