Matching

Prof. Mike Aguilar

https:/www.linkedin.com/in/mike-aguilar-econ

Matching - Preview

- Intuition: For each treated unit, find an untreated unit that "looks" similar. Subtract their outcomes. Average these differences for treatment effect.
- Exact matching: find a control that looks exactly like the treated
- Approximate matching: find a control that looks similar to the treated
- Who "looks alike" can depend on one or more covariates or probability of treatment (i.e. propensity scores)
- We can also weight the treat/control observations as an alternative to a simple DiM estimator

Housekeeping

- Throughout the slide deck you will see "Q", which indicates a question to you, the reader
- You will also see "A", which indicates the associated answer
- It is generally most efficient to learn this material through active participation. Whenever you encounter a "Q", be sure to try and develop your answer before turning the page to the provided "A" answer
- I recommend reading through these slides before engaging in associated coding exercises
- This slide deck is part of a broader curriculum in causal estimation and inference
- Suggested pre-requisite: Potential Outcomes, DAGs, control variables

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Case Study

How does military service impact salary?



Photo source: militarybases.com; Case study adapted from Angrist 1998

Background

- Veterans tend to earn more than civilians with similar backgrounds

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries
- Angrist 1998 examines the impact of volunteer military service in the US.

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries
- Angrist 1998 examines the impact of volunteer military service in the US.
- He tracks several hundred thousand men who voluntarily applied for military service in the late 1970s

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries
- Angrist 1998 examines the impact of volunteer military service in the US.
- He tracks several hundred thousand men who voluntarily applied for military service in the late 1970s
- For those that are accepted, they serve, are discharged, and then join civilian employment.

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries
- Angrist 1998 examines the impact of volunteer military service in the US.
- He tracks several hundred thousand men who voluntarily applied for military service in the late 1970s
- For those that are accepted, they serve, are discharged, and then join civilian employment.
- For those that are rejected, they move immediately to civilian employment

Background

- Veterans tend to earn more than civilians with similar backgrounds
- It is unclear whether there is a self-selection of the highest performers into the military or whether military service causes higher salaries
- Angrist 1998 examines the impact of volunteer military service in the US.
- He tracks several hundred thousand men who voluntarily applied for military service in the late 1970s
- For those that are accepted, they serve, are discharged, and then join civilian employment.
- For those that are rejected, they move immediately to civilian employment
- We will abstract away from the vagaries of military branch, MOS code (i.e. soldier's job), etc..

Data

- Unit of observation is the individual applicant (i)

Data

- Unit of observation is the individual applicant (i)
- Cross section is approximately $300,000$ individuals who voluntarily applied to the military between 1979 and 1982

Data

- Unit of observation is the individual applicant (i)
- Cross section is approximately $300,000$ individuals who voluntarily applied to the military between 1979 and 1982
- For each i we observe average annual salary in \$'s for civilian employment for the period 1988-1991

Data

- Unit of observation is the individual applicant (i)
- Cross section is approximately $300,000$ individuals who voluntarily applied to the military between 1979 and 1982
- For each i we observe average annual salary in \$'s for civilian employment for the period 1988-1991

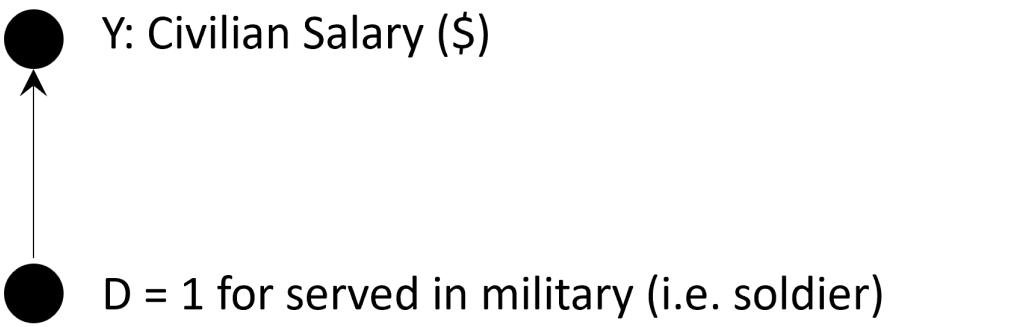
Causal Path

Q: Can you draw a DAG with only Outcome, Treatment, Treatment Effect?

Causal Path

Q: Can you draw a DAG with only Outcome, Treatment, Treatment Effect?

A:



- Outcome Y: Salary $ in 1988-1991
- Treatment D: Military service
- Treatment Effect: The impact of voluntary military service on salaries

Difference in Means

Q: Suppose avg(soldier income) minus avg(civilian income) = $\$18.2K$. Why might this naive Difference in Means not yield an appropriate TE estimate?

Difference in Means

Q: Suppose avg(soldier income) minus avg(civilian income) = $\$18.2K$. Why might this naive Difference in Means not yield an appropriate TE estimate?

A:

- $\widehat{ATE} = \bar{Y}_{Soldier} - \bar{Y}_{Civlian} = \$18.2K$ is likely plagued by selection bias.

Difference in Means

Q: Suppose avg(soldier income) minus avg(civilian income) = $\$18.2K$. Why might this naive Difference in Means not yield an appropriate TE estimate?

A:

- $\widehat{ATE} = \bar{Y}_{Soldier} - \bar{Y}_{Civlian} = \$18.2K$ is likely plagued by selection bias.
- Soldiers might be inherently different than civilians.

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

A:

- $\bar{Y}_{Soldier}$ is our proxy for $E[Y_{1i}|D_i = 1]$, ie the potential salary outcome for accepted applicants that served in the military

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

A:

- $\bar{Y}_{Soldier}$ is our proxy for $E[Y_{1i}|D_i = 1]$, ie the potential salary outcome for accepted applicants that served in the military
- $\bar{Y}_{Civilian}$ is our proxy for $E[Y_{0i}|D_i = 0]$, ie the potential salary outcome for rejected applicants that did not serve in the military

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

A:

- $\bar{Y}_{Soldier}$ is our proxy for $E[Y_{1i}|D_i = 1]$, ie the potential salary outcome for accepted applicants that served in the military
- $\bar{Y}_{Civilian}$ is our proxy for $E[Y_{0i}|D_i = 0]$, ie the potential salary outcome for rejected applicants that did not serve in the military
- By taking the simple difference in means our estimate is capturing two forces

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

A:

- $\bar{Y}_{Soldier}$ is our proxy for $E[Y_{1i}|D_i = 1]$, ie the potential salary outcome for accepted applicants that served in the military
- $\bar{Y}_{Civilian}$ is our proxy for $E[Y_{0i}|D_i = 0]$, ie the potential salary outcome for rejected applicants that did not serve in the military
- By taking the simple difference in means our estimate is capturing two forces

$$\begin{aligned}(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]) =& (E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]) \\ &+ (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])\end{aligned}$$

$$\bar{Y}_{Soldier} - \bar{Y}_{Civilian} \approx$$

Potential Outcomes

Q: Can you relate the difference in means to the potential outcomes framework?

A:

- $\bar{Y}_{Soldier}$ is our proxy for $E[Y_{1i}|D_i = 1]$, ie the potential salary outcome for accepted applicants that served in the military
- $\bar{Y}_{Civilian}$ is our proxy for $E[Y_{0i}|D_i = 0]$, ie the potential salary outcome for rejected applicants that did not serve in the military
- By taking the simple difference in means our estimate is capturing two forces

$$\begin{aligned}(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]) =& (E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]) \\ &+ (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])\end{aligned}$$

$$\bar{Y}_{Soldier} - \bar{Y}_{Civilian} \approx ATT + \text{Selection Bias}$$

Interpreting the Selection Bias

Q: Can you interpret the Selection Bias in our case?

$$\text{Selection Bias} = (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

Interpreting the Selection Bias

Q: Can you interpret the Selection Bias in our case?

$$\text{Selection Bias} = (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

A:

- Selection bias is the income soldiers would have earned if they didn't serve in the military minus the salary of civilians

Interpreting the Selection Bias

Q: Can you interpret the Selection Bias in our case?

$$\text{Selection Bias} = (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

A:

- Selection bias is the income soldiers would have earned if they didn't serve in the military minus the salary of civilians
- The case study background suggests this might be positive
- The selection bias captures inherent differences between applicants who served and those that didn't. Perhaps this is a measure of natural ability.

Interpreting the Selection Bias

Q: Can you interpret the Selection Bias in our case?

$$\text{Selection Bias} = (E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])$$

A:

- Selection bias is the income soldiers would have earned if they didn't serve in the military minus the salary of civilians
- The case study background suggests this might be positive
- The selection bias captures inherent differences between applicants who served and those that didn't. Perhaps this is a measure of natural ability.
- As part of the application process, the US military mandates an aptitude test. Higher scores increases the chances of acceptance.
- We will assume that this test is a suitable proxy for natural ability.

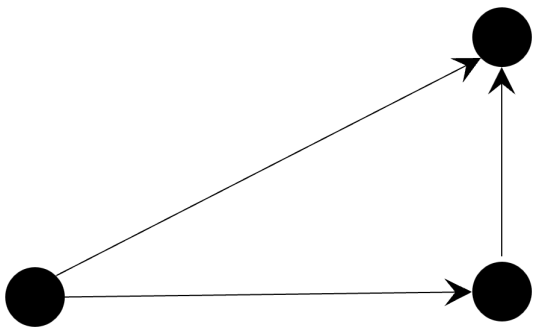
Aptitude Test

Q: Can you draw the DAG with aptitude?

Aptitude Test

Q: Can you draw the DAG with aptitude?

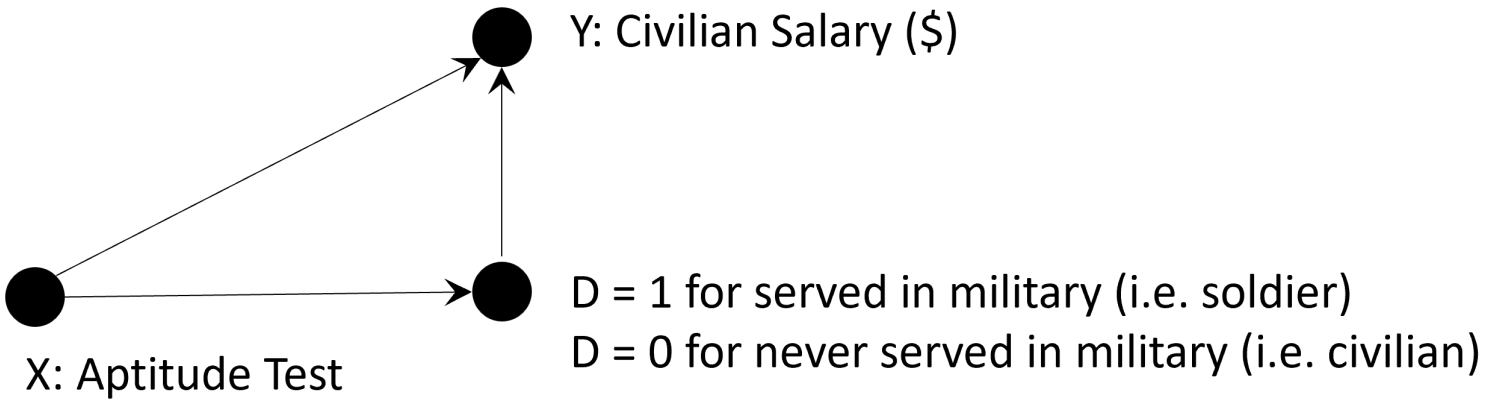
A:



Aptitude Test

Q: Can you draw the DAG with aptitude?

A:

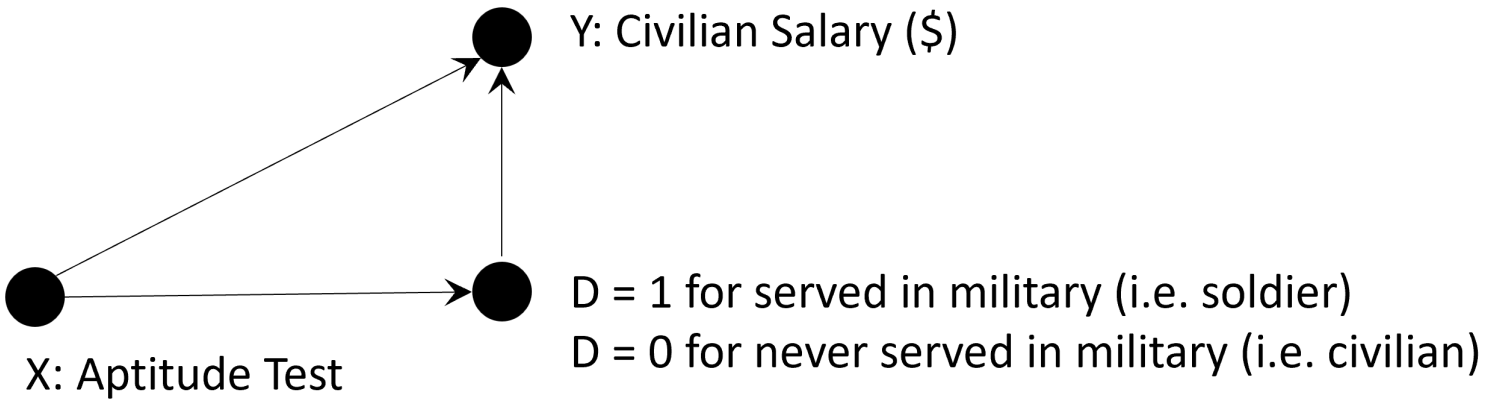


Q: Why is the aptitude test score a "Good" control?

Aptitude Test

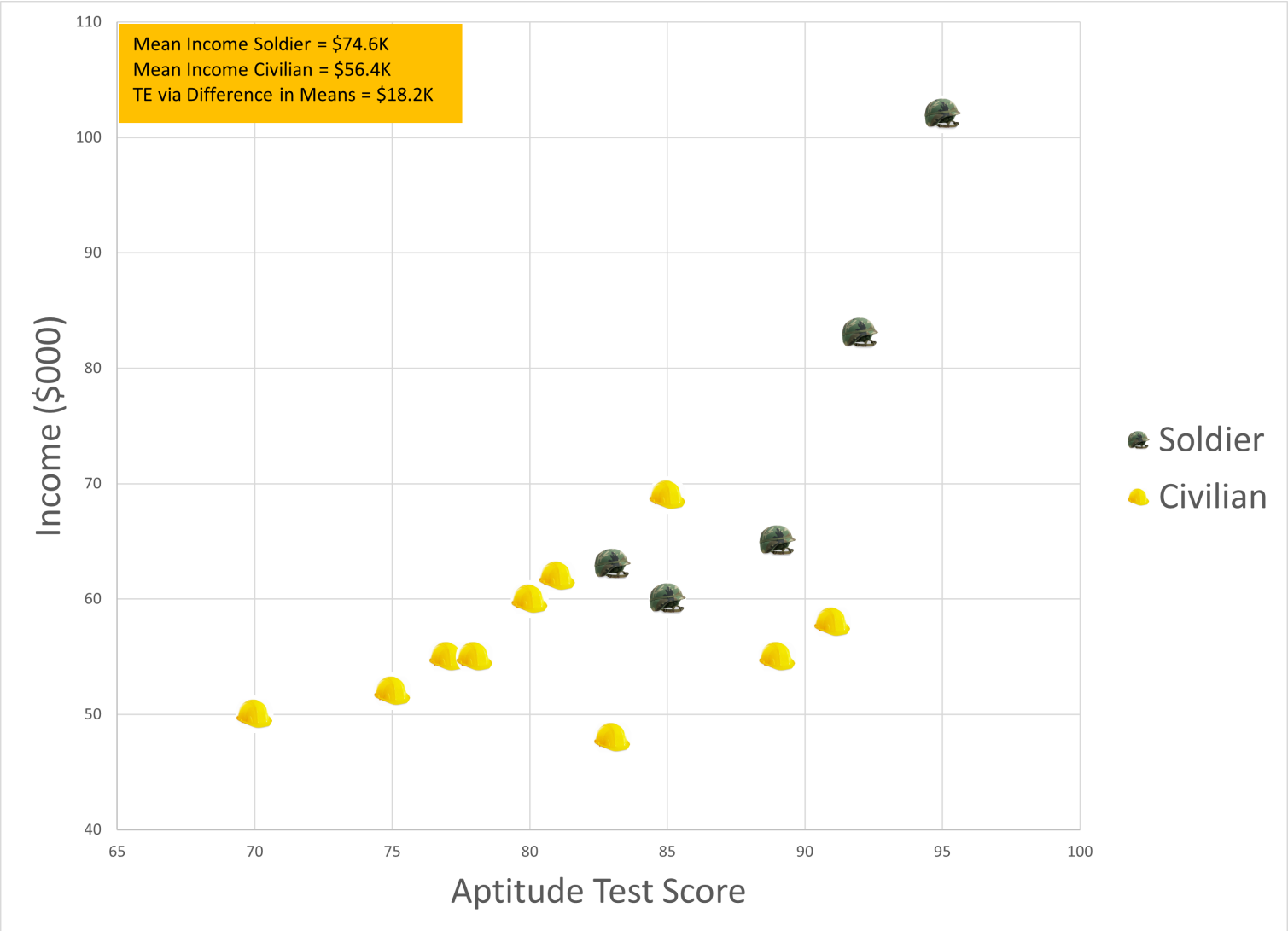
Q: Can you draw the DAG with aptitude?

A:



Q: Why is the aptitude test score a "Good" control?
A: The aptitude test score is related to the treatment D (i.e. those with higher aptitudes are more likely to be accepted), and the outcome variable Y directly (i.e. those with higher natural aptitudes might earn higher salaries regardless of military service); i.e. aptitude test score is a confounder.

Peaking at our Data



Non Compliers

Q: Who might be a "non complier"? Why might they prove troublesome for the analysis?

Non Compliers

Q: Who might be a "non complier"? Why might they prove troublesome for the analysis?

A:

- A non-complier would be someone who applied, was accepted, but decided not to enlist

Non Compliers

Q: Who might be a "non complier"? Why might they prove troublesome for the analysis?

A:

- A non-complier would be someone who applied, was accepted, but decided not to enlist
- In our case, compliance means that if someone applies and is accepted into the military, they serve...
- ...and if someone is rejected, they cannot serve

Non Compliers

Q: Who might be a "non complier"? Why might they prove troublesome for the analysis?

A:

- A non-complier would be someone who applied, was accepted, but decided not to enlist
- In our case, compliance means that if someone applies and is accepted into the military, they serve...
- ...and if someone is rejected, they cannot serve
- Non-compliers obscure our link between ability (aptitude test) and military service

Non Compliers

Q: Who might be a "non complier"? Why might they prove troublesome for the analysis?

A:

- A non-complier would be someone who applied, was accepted, but decided not to enlist
- In our case, compliance means that if someone applies and is accepted into the military, they serve...
- ...and if someone is rejected, they cannot serve
- Non-compliers obscure our link between ability (aptitude test) and military service
- We can use IV (Instrumental Variable) methods to deal with non-compliance, but that is beyond our scope here
- For ease of exposition, we assume full compliance

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores
- How measure "similar"?: i) (Normalized) Euclidean Distance, ii) Mahalanobis

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores
- How measure "similar"?: i) (Normalized) Euclidean Distance, ii) Mahalanobis
- How form clusters of similar individuals (aka matches)?: i) Exact Matching, ii) Approximate Matching (k-NN, Subclassification, Weighting)

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores
- How measure "similar"?: i) (Normalized) Euclidean Distance, ii) Mahalanobis
- How form clusters of similar individuals (aka matches)?: i) Exact Matching, ii) Approximate Matching (k-NN, Subclassification, Weighting)
- How assess/ensure high quality matches?: i) Diagnostics, ii) Caliper, iii) Bias Correction

Overview

Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores
- How measure "similar"?: i) (Normalized) Euclidean Distance, ii) Mahalanobis
- How form clusters of similar individuals (aka matches)?: i) Exact Matching, ii) Approximate Matching (k-NN, Subclassification, Weighting)
- How assess/ensure high quality matches?: i) Diagnostics, ii) Caliper, iii) Bias Correction
- Which Treatment Effect?: i) ATE, ii) ATT, iii) ATC, iv) CATE

Overview

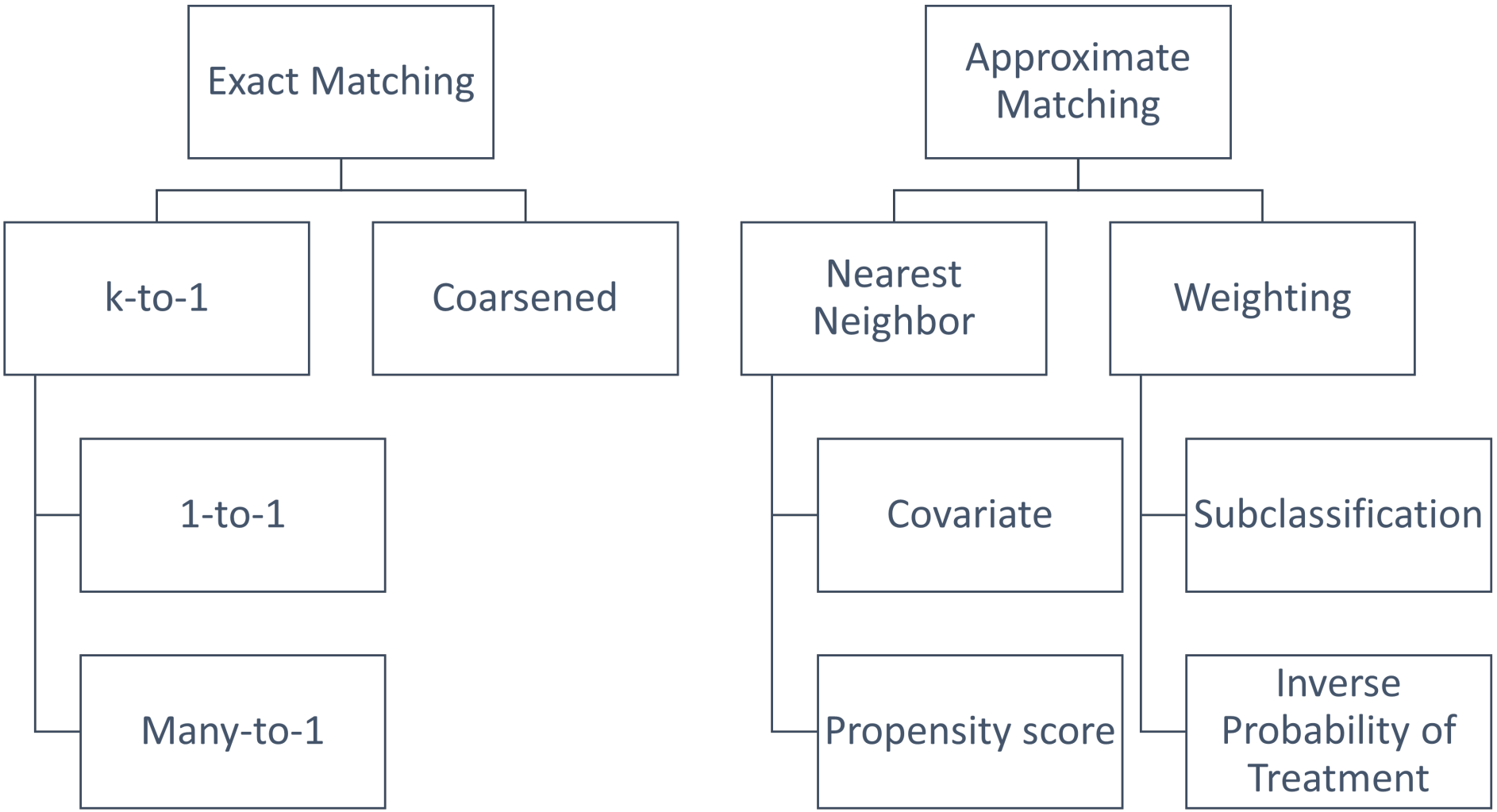
Main Idea: For each soldier find a similar looking civilian (aka match). Compare their salaries.
Here is a (non exhaustive) list of considerations when computing TE via matching:

- How are you defining "similar"?: i) Covariates, 2) Propensity Scores
- How measure "similar"?: i) (Normalized) Euclidean Distance, ii) Mahalanobis
- How form clusters of similar individuals (aka matches)?: i) Exact Matching, ii) Approximate Matching (k-NN, Subclassification, Weighting)
- How assess/ensure high quality matches?: i) Diagnostics, ii) Caliper, iii) Bias Correction
- Which Treatment Effect?: i) ATE, ii) ATT, iii) ATC, iv) CATE
- How estimate?: i) Difference in Means, ii) Regression Adjustment

Hierarchy of Causal Settings

1. Randomized Control Trials
2. Natural Experiments
3. Observational Studies
    - Matching is an observational study

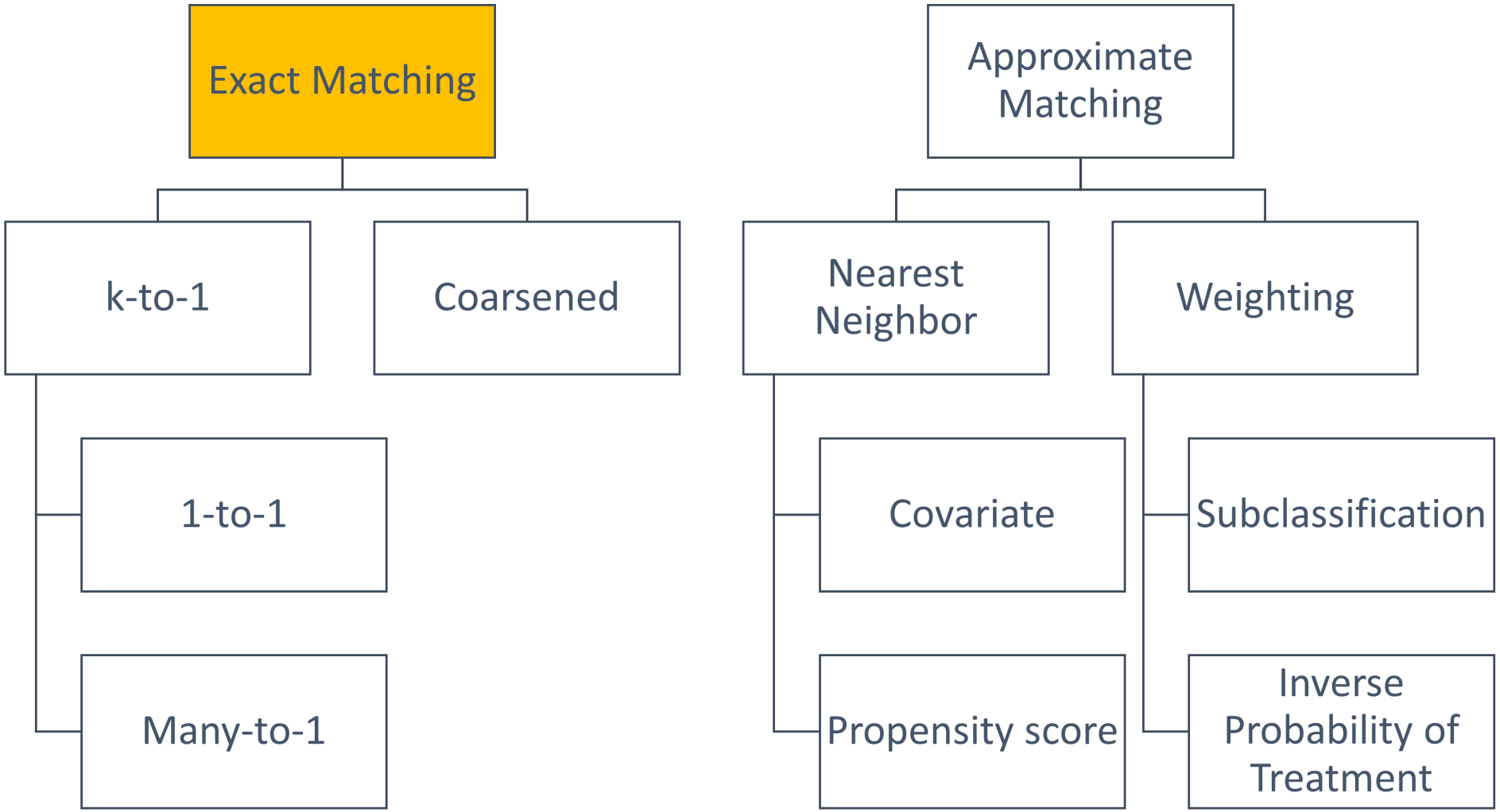
Hierarchy of (some) Matching Methods



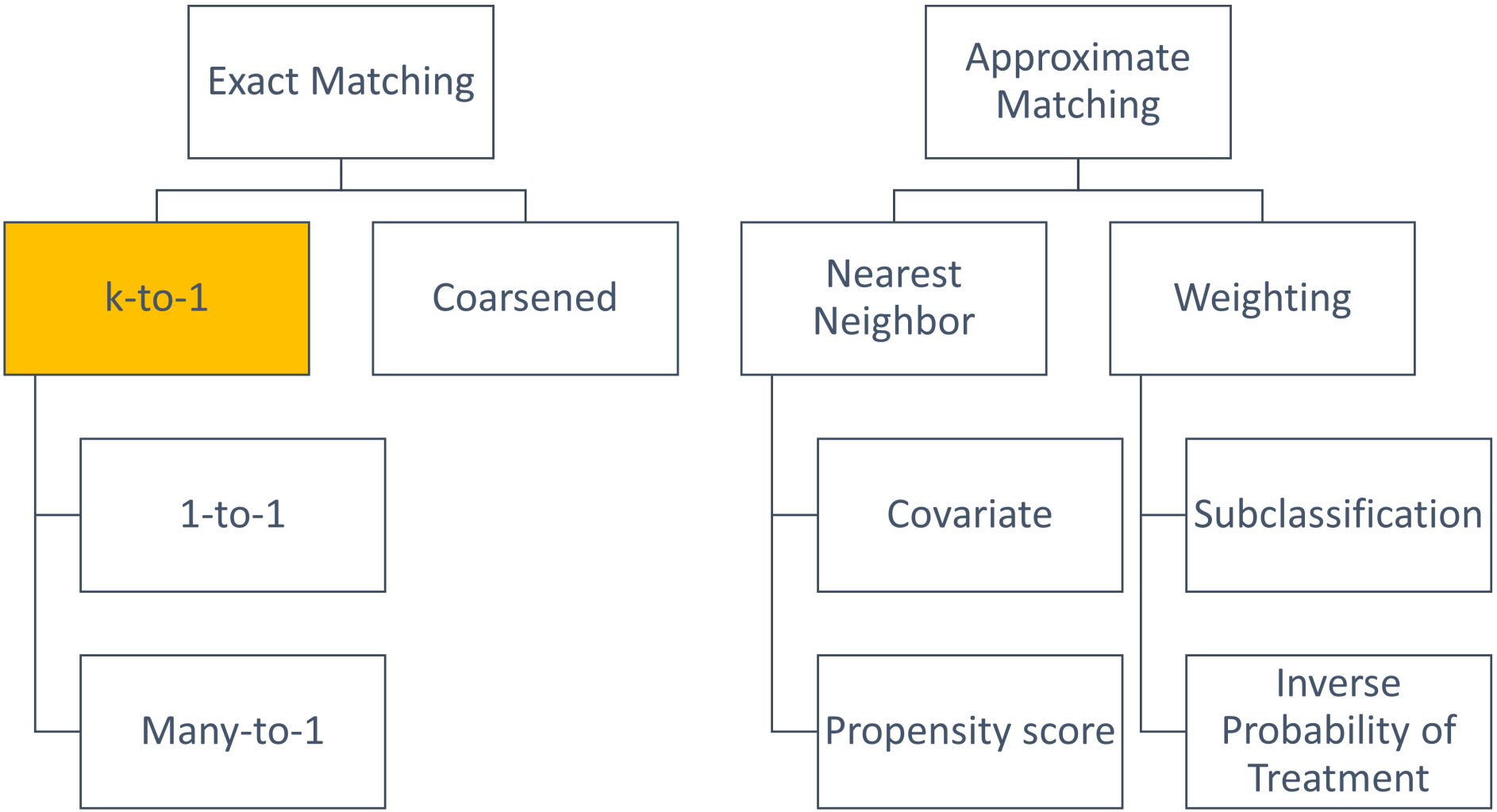
Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Exact Matching



Exact Matching k-to-1

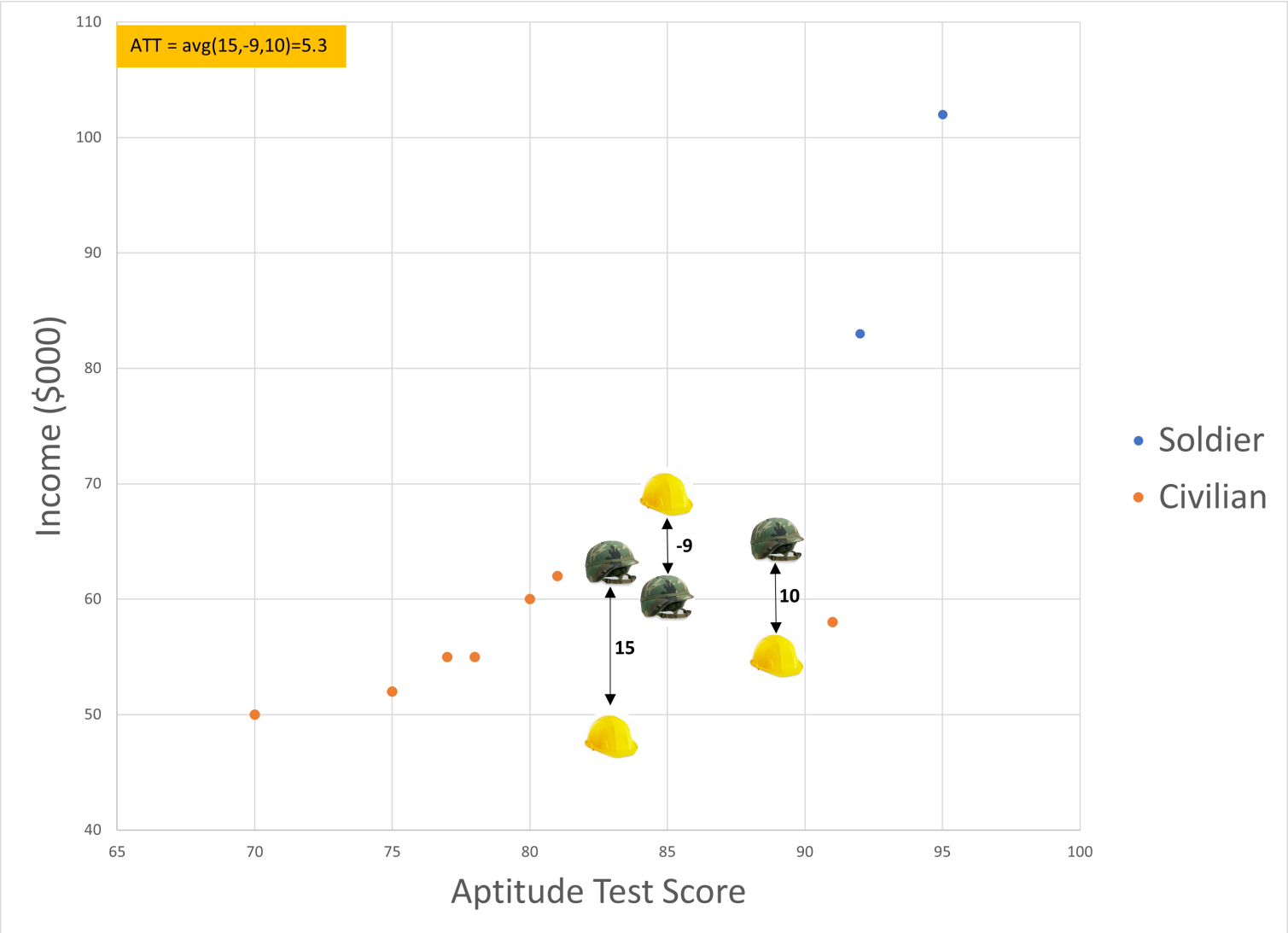


Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Exact Matching (1-to-1)

ATT: For each treated, find one matching control. Subtract incomes and compute avg.



Strictly speaking, if some treated units are dropped due to lack of corresponding control, the estimand is no longer the ATT.
Exact 1-to-1 matching can be conducted on multiple covariates.

Exact Matching (1-to-1)

With 1-to-1 Exact Matching ATE=ATT=ATC

Exact Matching (1-to-1)

With 1-to-1 Exact Matching ATE=ATT=ATC

ID	D(Soldier=1)	Test Score	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	63	48	15	1	15
2	1	85	60	69	-9	1	-9
3	1	89	65	55	10	1	10
4	1	92	83			1	
5	1	95	102			1	
6	0	70	50			-1	
7	0	75	52			-1	
8	0	77	55			-1	
9	0	78	55			-1	
10	0	80	60			-1	
11	0	81	62			-1	
12	0	83	48	63	-15	-1	15
13	0	85	69	60	9	-1	-9
14	0	89	55	65	-10	-1	10
15	0	91	58			-1	

Matched Units (Treated or Control) = 6 | ATE = 5.3 || # Matched Treated Units = 3 | ATT = 5.3 || # Matched Control Units = 3 | ATC = 5.3

In this graphic, the colors represent "matched" observations. For instance, IDs 1 and 12 are both red since they are matched on their Test Score.

Exact Matching (1-to-1)

The Role of CATE and 2D-1

- Q: What's the role of $2D - 1$?

Exact Matching (1-to-1)

The Role of CATE and 2D-1

- Q: What's the role of $2D - 1$?
- A:
  - In the prior table, if we blindly apply the formula of Income Minus Counterfactual Income across all the rows the CATE for the control units would be negative.

Exact Matching (1-to-1)

The Role of CATE and 2D-1

- Q: What's the role of $2D - 1$?
- A:
  - In the prior table, if we blindly apply the formula of Income Minus Counterfactual Income across all the rows the CATE for the control units would be negative.
  - In our case study that would imply that being accepted into the military would reduce incomes for the civilians, which is the opposite of our intuition and the reverse of the potential outcomes framework that computes treated state minus control state.

Exact Matching (1-to-1)

The Role of CATE and 2D-1

- Q: What's the role of $2D - 1$?
- A:
  - In the prior table, if we blindly apply the formula of Income Minus Counterfactual Income across all the rows the CATE for the control units would be negative.
  - In our case study that would imply that being accepted into the military would reduce incomes for the civilians, which is the opposite of our intuition and the reverse of the potential outcomes framework that computes treated state minus control state.
  - $2D - 1$ is a numerical trick to resolve the situation.

Exact Matching (1-to-1)

The Role of CATE and 2D-1

- Q: What's the role of $2D - 1$?
- A:
  - In the prior table, if we blindly apply the formula of Income Minus Counterfactual Income across all the rows the CATE for the control units would be negative.
  - In our case study that would imply that being accepted into the military would reduce incomes for the civilians, which is the opposite of our intuition and the reverse of the potential outcomes framework that computes treated state minus control state.
  - $2D - 1$ is a numerical trick to resolve the situation.
  - When $D = 1$ for treated, $2D - 1 = 1$, which gives a positive sign to observed income of the soldiers minus the unobserved counterfactual income of the civilians to whom they are matched.

Exact Matching (1-to-1)

The Role of CATE and 2D-1

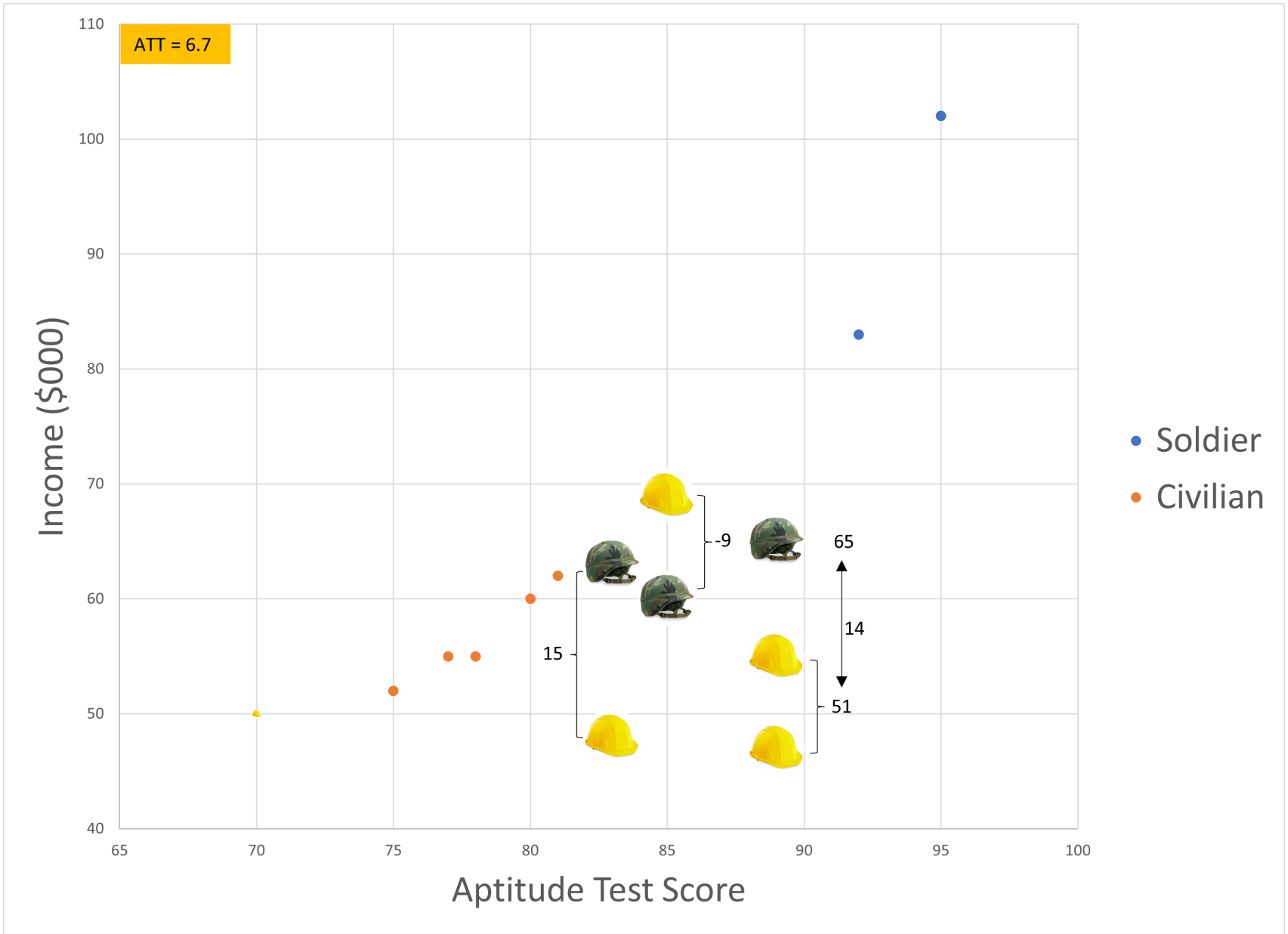
- Q: What's the role of $2D - 1$?
- A:
  - In the prior table, if we blindly apply the formula of Income Minus Counterfactual Income across all the rows the CATE for the control units would be negative.
  - In our case study that would imply that being accepted into the military would reduce incomes for the civilians, which is the opposite of our intuition and the reverse of the potential outcomes framework that computes treated state minus control state.
  - $2D - 1$ is a numerical trick to resolve the situation.
  - When $D = 1$ for treated, $2D - 1 = 1$, which gives a positive sign to observed income of the soldiers minus the unobserved counterfactual income of the civilians to whom they are matched.
  - When $D = 0$ for control, $2D - 1 = -1$, which gives a negative sign to the observed income of the civilians minus the unobserved counterfactual income of the soldiers to whom they are matched.

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

many-to-1

Avg income diff. across the potentially "many" matching Civilians for each "1" Soldier



many-to-1

It is common to have more controls than treated due to cost of treatment, etc..

many-to-1

It is common to have more controls than treated due to cost of treatment, etc..

If there are multiple controls for a given treated, then we can average among the controls to serve as counterfactual. Here we are averaging only across the number of matched treated units:

$$\widehat{ATT} = \frac{1}{N^{MatchedSoldiers}} \sum_{MatchedSoliders} (Y_{1i}^{Soldier} - avg(\tilde{Y}_{0i}^{Soldier}))$$

where $(\tilde{Y}_{0i}^{Soldier} = Y_{0i}^{Civilian}) | AptScore_i^{Soldier} = AptScore_i^{Civilian}$

many-to-1

It is common to have more controls than treated due to cost of treatment, etc..

If there are multiple controls for a given treated, then we can average among the controls to serve as counterfactual. Here we are averaging only across the number of matched treated units:

$$\widehat{ATT} = \frac{1}{N^{MatchedSoldiers}} \sum_{MatchedSoliders} (Y_{1i}^{Soldier} - avg(\tilde{Y}_{0i}^{Soldier}))$$

where $(\tilde{Y}_{0i}^{Soldier} = Y_{0i}^{Civilian}) | AptScore_i^{Soldier} = AptScore_i^{Civilian}$

- If multiple soldiers have the same aptitude score we focus on each soldier separately, and either i) uniquely assign civilians as matches, or ii) impose the same/similar group of civilians to each soldier
- If soldiers are discarded in the matching process, then this estimate is no longer technically the ATT

many-to-1

Q: How would you compute ATC and ATE?

many-to-1

Q: How would you compute ATC and ATE?

A:

- ATC would be computed analogously, using the matched soldiers as the counterfactual for each matched civilian

many-to-1

Q: How would you compute ATC and ATE?

A:

- ATC would be computed analogously, using the matched soldiers as the counterfactual for each matched civilian
- ATE would be average over over all matched units

many-to-1

Q: How would you compute ATC and ATE?

A:

- ATC would be computed analogously, using the matched soldiers as the counterfactual for each matched civilian
- ATE would be average over over all matched units
- Often, g-computation is used, wherein every unit in the sample is assigned a counterfactual value via the matching process. Then we average across the desired subset (treated, control, all).

many-to-1

ID	D(Soldier=1)	Test Score	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	63	48	15	1	15
2	1	85	60	69	-9	1	-9
3	1	89	65	51	14	1	14
4	1	92	83				
5	1	95	102				
6	0	70	50				
7	0	75	52				
8	0	77	55				
9	0	78	55				
10	0	80	60				
11	0	81	62				
12	0	83	48	63	-15	-1	15
13	0	85	69	60	9	-1	-9
14	0	89	55	65	-10	-1	10
15	0	89	47	65	-18	-1	18
# Matched Units (Treated or Control) = 7 | ATE = 7.7 || # Matched Treated Units = 3 | ATT = 6.7 || # Matched Control Units = 4 | ATC = 8.5							

Note: many-to-1 is the default (and only) option in popular software implementations (e.g. MatchIt) of Exact Matching. The concept of 1-to-1 and many-to-1 applies to Approximate Matching as well.

Matching over multiple dimensions

Q: What's the (dis)advantage of matching by Test Score AND Schooling?

Matching over multiple dimensions

Q: What's the (dis)advantage of matching by Test Score AND Schooling?

A:

ID	D(Soldier=1)	Test Score	Yrs School Completed	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	12	63				
2	1	85	14	60	69	-9	1	-9
3	1	89	16	65				
4	1	92	16	83				
5	1	95	16	102				
6	0	70	11	50				
7	0	75	12	52				
8	0	77	12	55				
9	0	78	12	55				
10	0	80	12	60				
11	0	81	12	62				
12	0	83	11	48				
13	0	85	14	69	60	9	-1	-9
14	0	89	14	55				
15	0	89	14	47				

Matching over multiple dimensions

Q: What's the (dis)advantage of matching by Test Score AND Schooling?

A:

ID	D(Soldier=1)	Test Score	Yrs School Completed	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	12	63				
2	1	85	14	60	69	-9	1	-9
3	1	89	16	65				
4	1	92	16	83				
5	1	95	16	102				
6	0	70	11	50				
7	0	75	12	52				
8	0	77	12	55				
9	0	78	12	55				
10	0	80	12	60				
11	0	81	12	62				
12	0	83	11	48				
13	0	85	14	69	60	9	-1	-9
14	0	89	14	55				
15	0	89	14	47				

- Advantage: better matches

Matching over multiple dimensions

Q: What's the (dis)advantage of matching by Test Score AND Schooling?

A:

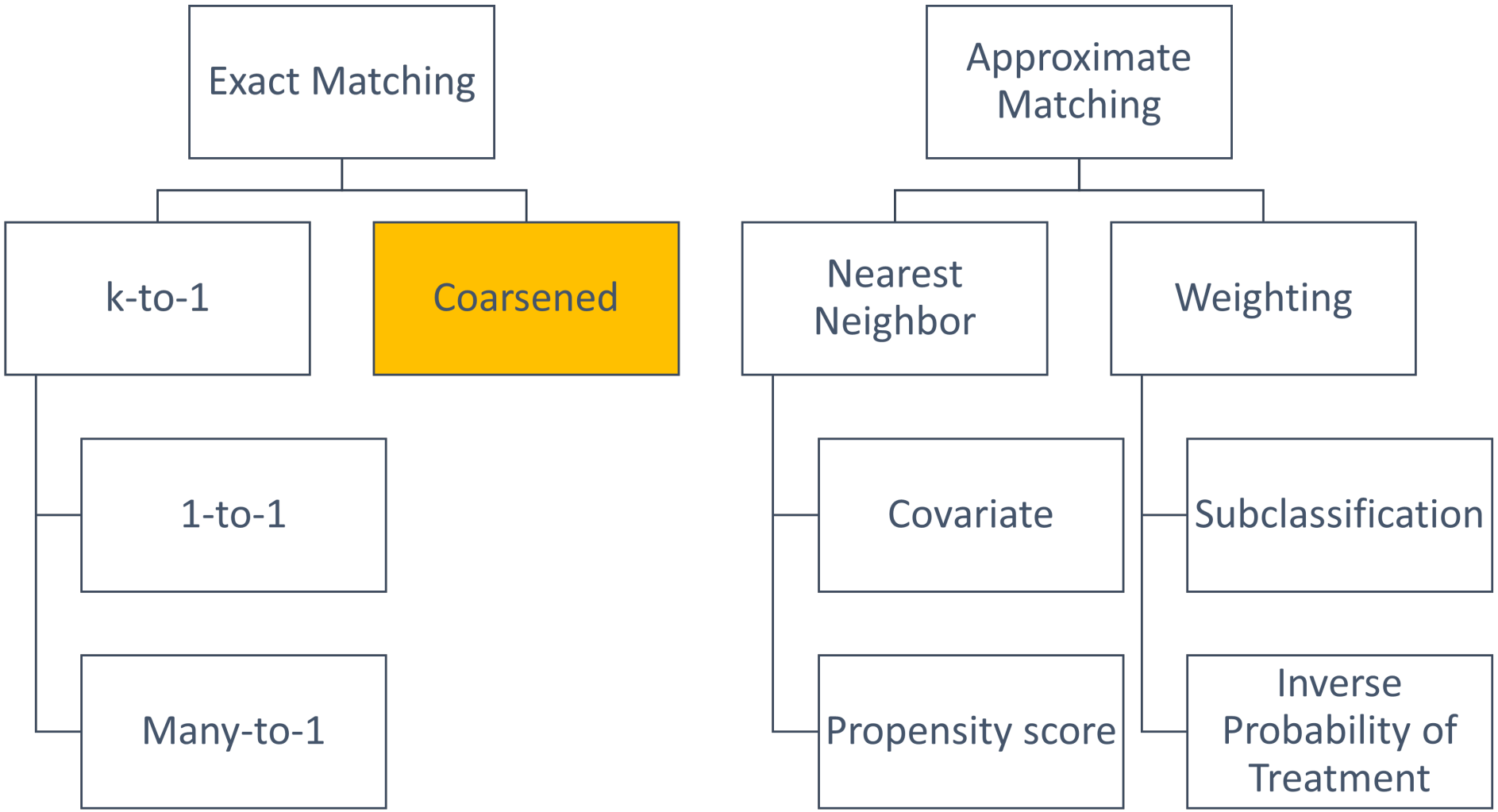
ID	D(Soldier=1)	Test Score	Yrs School Completed	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	12	63				
2	1	85	14	60	69	-9	1	-9
3	1	89	16	65				
4	1	92	16	83				
5	1	95	16	102				
6	0	70	11	50				
7	0	75	12	52				
8	0	77	12	55				
9	0	78	12	55				
10	0	80	12	60				
11	0	81	12	62				
12	0	83	11	48				
13	0	85	14	69	60	9	-1	-9
14	0	89	14	55				
15	0	89	14	47				

- Advantage: better matches
- Disadvantage: lose many observations

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Coarsened Exact Matching



Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling
3. Use the matched Civilian as a counterfactual for the Soldier's income

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling
3. Use the matched Civilian as a counterfactual for the Soldier's income
4. Compute CATE for that matched unit and multiply by $2D - 1$ to adjust for sign

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling
3. Use the matched Civilian as a counterfactual for the Soldier's income
4. Compute CATE for that matched unit and multiply by $2D - 1$ to adjust for sign
5. Average the adjusted CATE's across the number of treated units for the ATT

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling
3. Use the matched Civilian as a counterfactual for the Soldier's income
4. Compute CATE for that matched unit and multiply by $2D - 1$ to adjust for sign
5. Average the adjusted CATE's across the number of treated units for the ATT

- Adapt the algorithm above for ATE/ATC by finding the counterfactual income for each Civilian among the Soldiers. Average the CATEs across the controls for ATC and across all units for the ATE

Coarsened Exact Matching

Coarsening adjusts Exact Matching; Let's start with the algo, then discuss intuition

1. Coarsen Test Scores into grade deciles
2. For each Soldier find a Civilian with the same test score decile AND years of schooling
3. Use the matched Civilian as a counterfactual for the Soldier's income
4. Compute CATE for that matched unit and multiply by $2D - 1$ to adjust for sign
5. Average the adjusted CATE's across the number of treated units for the ATT

- Adapt the algorithm above for ATE/ATC by finding the counterfactual income for each Civilian among the Soldiers. Average the CATEs across the controls for ATC and across all units for the ATE
- Coarsening usually done via subject matter expertise (e.g. scoring in the 80s, 90s, etc.. makes sense in our case). Automatic coarsening algorithms are available

Coarsened Exact Matching Example

ID	D(Soldier=1)	Test Score	Test Score Decile	Yrs School Completed	Income	Counterfactual Income	CATE	2D-1	CATE*(2D-1)
1	1	83	80-89	12	63	61	2	1	2.0
2	1	85	80-89	14	60	55	5	1	5.3
3	1	89	80-89	16	65				
4	1	92	90-100	16	83				
5	1	95	90-100	16	102				
6	0	70	70-79	11	50				
7	0	75	70-79	12	52				
8	0	77	70-79	12	55				
9	0	78	70-79	12	55				
10	0	80	80-89	12	60	63	-3	-1	3.0
11	0	81	80-89	12	62	63	-1	-1	1.0
12	0	83	80-89	11	48				
13	0	85	80-89	14	69	60	9	-1	-9.0
14	0	89	80-89	14	55	60	-5	-1	5.0
15	0	89	80-89	14	40	60	-20	-1	20.0

Matched Units (Treated or Control) = 7 | ATE = 3.9 || # Matched Treated Units = 2 | ATT = 3.7 || # Matched Control Units = 5 | ATC = 4.0

(Dis)Advantages of Coarsened Exact Matching

Q: What are the (dis)advantages of coarsening for matching?

(Dis)Advantages of Coarsened Exact Matching

Q: What are the (dis)advantages of coarsening for matching?

A:

Advantages:

- Coarsening generates more matches, which can improve precision

(Dis)Advantages of Coarsened Exact Matching

Q: What are the (dis)advantages of coarsening for matching?

A:

Advantages:

- Coarsening generates more matches, which can improve precision

Disadvantages:

- More coarsening leads to covariate imbalance, which can lead to more bias

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

(Dis)advantages of Exact Matching

Q: What are the (dis)advantages of Exact Matching?

(Dis)advantages of Exact Matching

Q: What are the (dis)advantages of Exact Matching?

A:

Advantages

- Exact matches increase the chances of mitigating selection bias

(Dis)advantages of Exact Matching

Q: What are the (dis)advantages of Exact Matching?

A:

Advantages

- Exact matches increase the chances of mitigating selection bias

Disadvantages

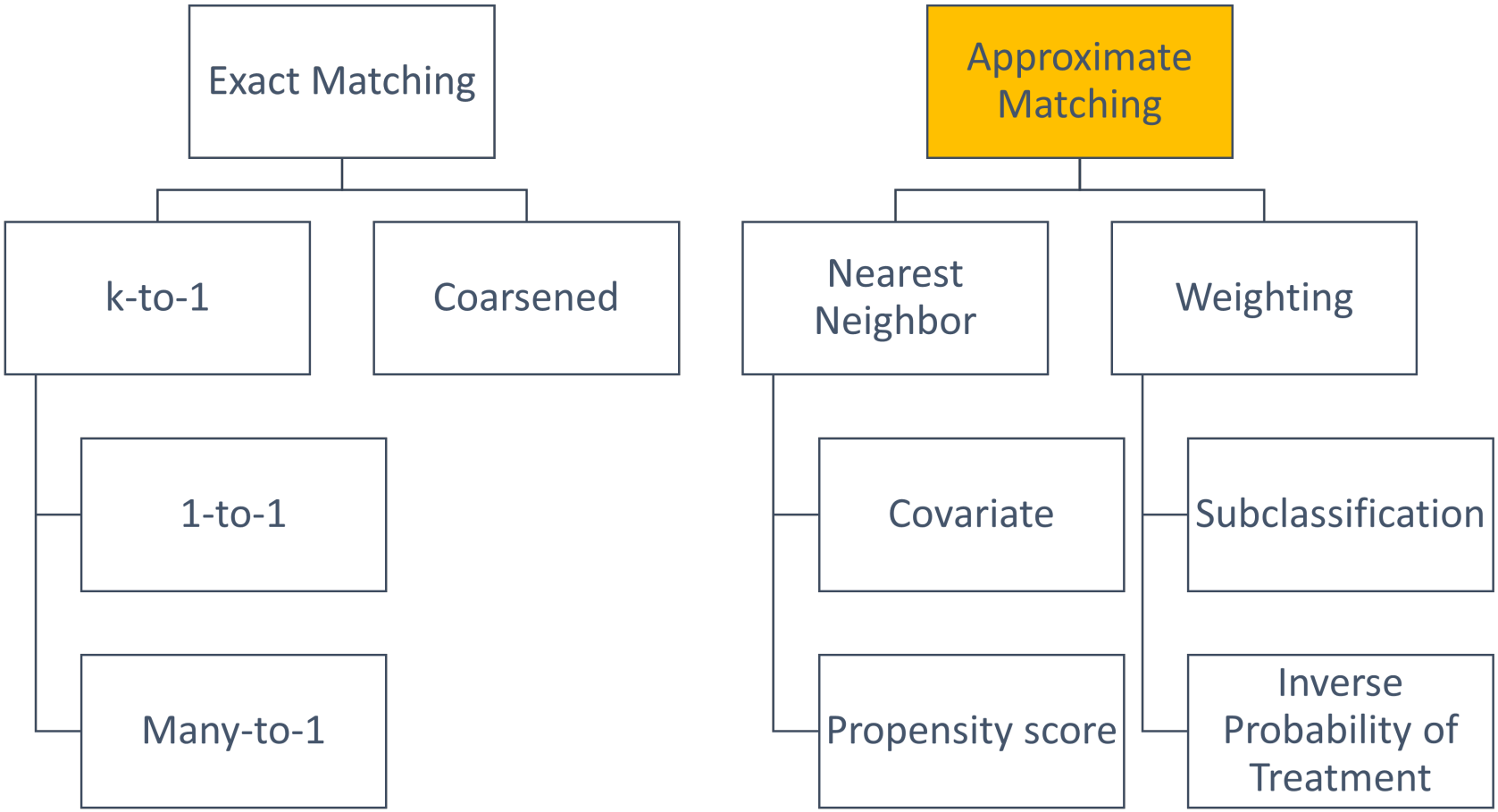
- Potential to lose many observations grows with more covariates / finer granularity of covariates; fewer observations (generally) implies less precision

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

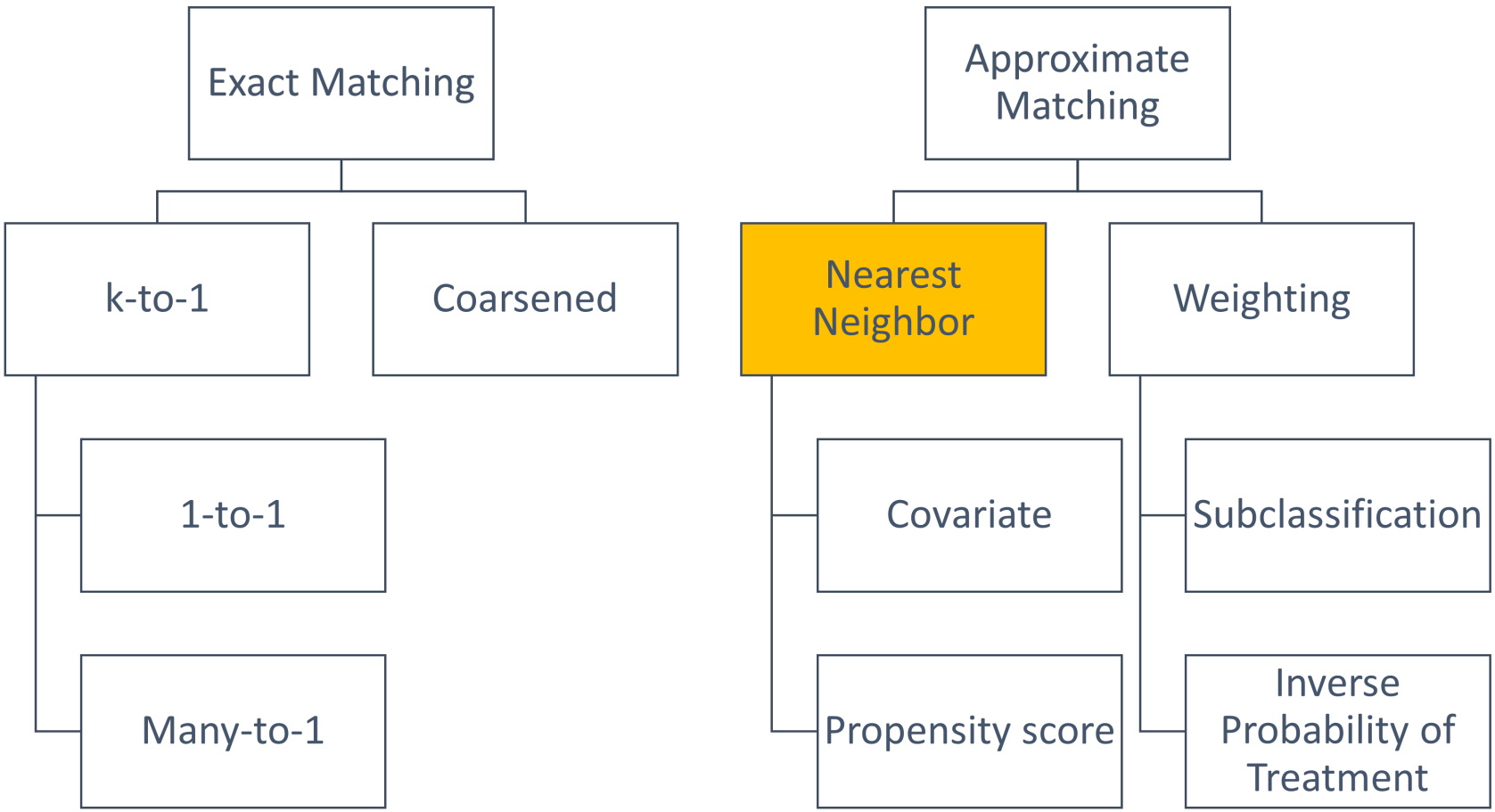
Approximate Matching

Anytime we approximate there are likely to be "errors"



Nearest Neighbor Matching

Can find matches via Covariates or Propensity Scores



Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)
2. Find the k Civilians (control units) most similar to Soldier (aka matches)

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)
2. Find the k Civilians (control units) most similar to Soldier (aka matches)
3. Average Salary (outcome) across matched Civilians

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)
2. Find the k Civilians (control units) most similar to Soldier (aka matches)
3. Average Salary (outcome) across matched Civilians
4. Use that average as counterfactual for Soldier

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)
2. Find the k Civilians (control units) most similar to Soldier (aka matches)
3. Average Salary (outcome) across matched Civilians
4. Use that average as counterfactual for Soldier
5. Repeat for all Soldiers

Nearest Neighbor Matching Algo

Prelims: i) distance metric, ii) order of treated, iii) $k = \#$ controls per treated, iv) w(out) replacement, v) object to match (covariates or propensity score)

1. Pick a Soldier (treated unit)
2. Find the k Civilians (control units) most similar to Soldier (aka matches)
3. Average Salary (outcome) across matched Civilians
4. Use that average as counterfactual for Soldier
5. Repeat for all Soldiers

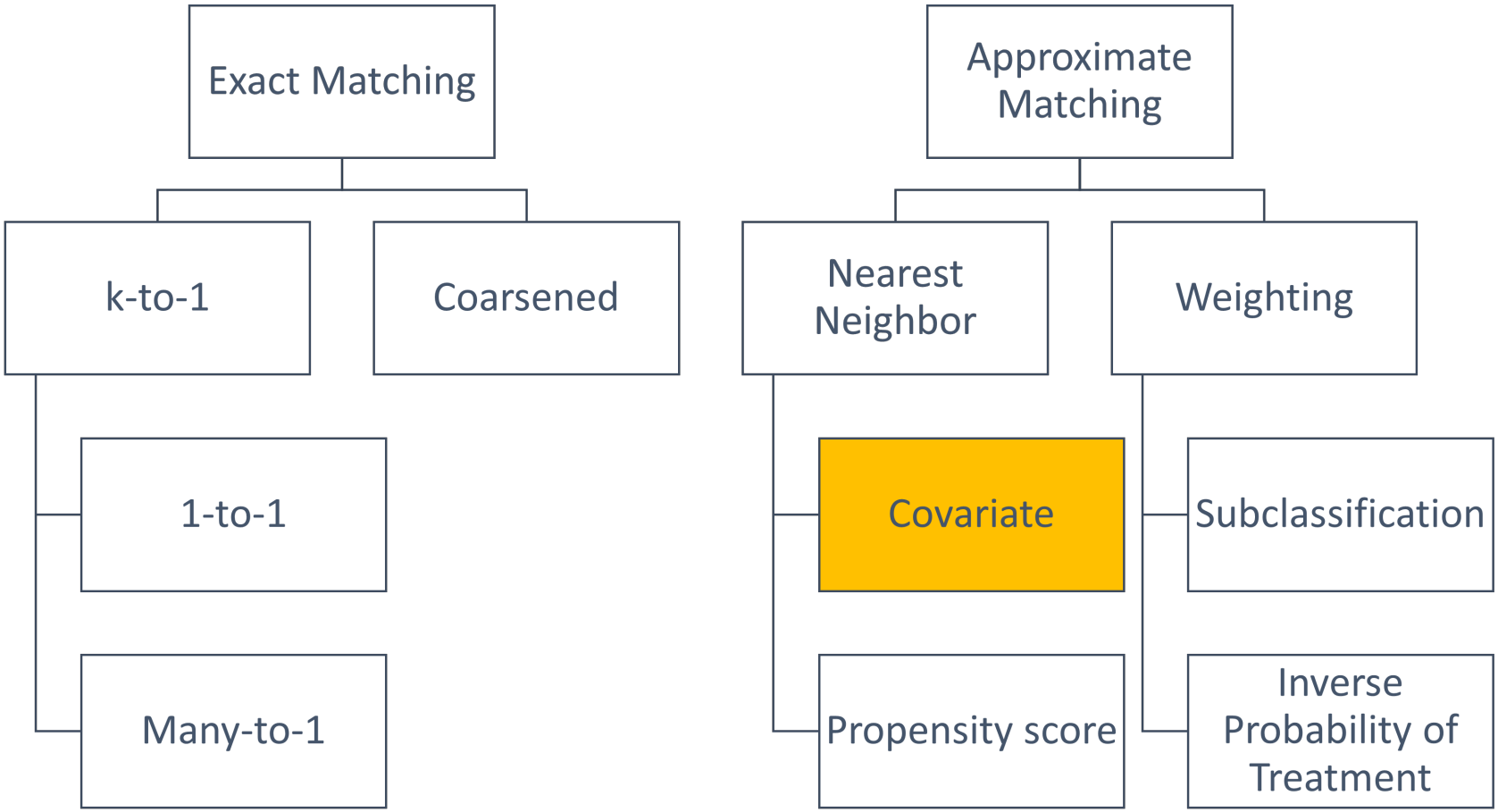
Note: Although technically distinct, this Nearest Neighbor Matching is similar to kNN. They both use a distance measure to identify units that are similar to one another. They are both supervised in the sense that we direct the dimension of similarity (i.e. Covariates or Propensity scores). In standard kNN there is uncertainty over the labels, which is why we use Test and Training sets to evaluate the accuracy of matches. In our k:1 Nearest Neighbor Matching there is no such uncertainty; Soldiers are treated and Civilians are controls.

Outline

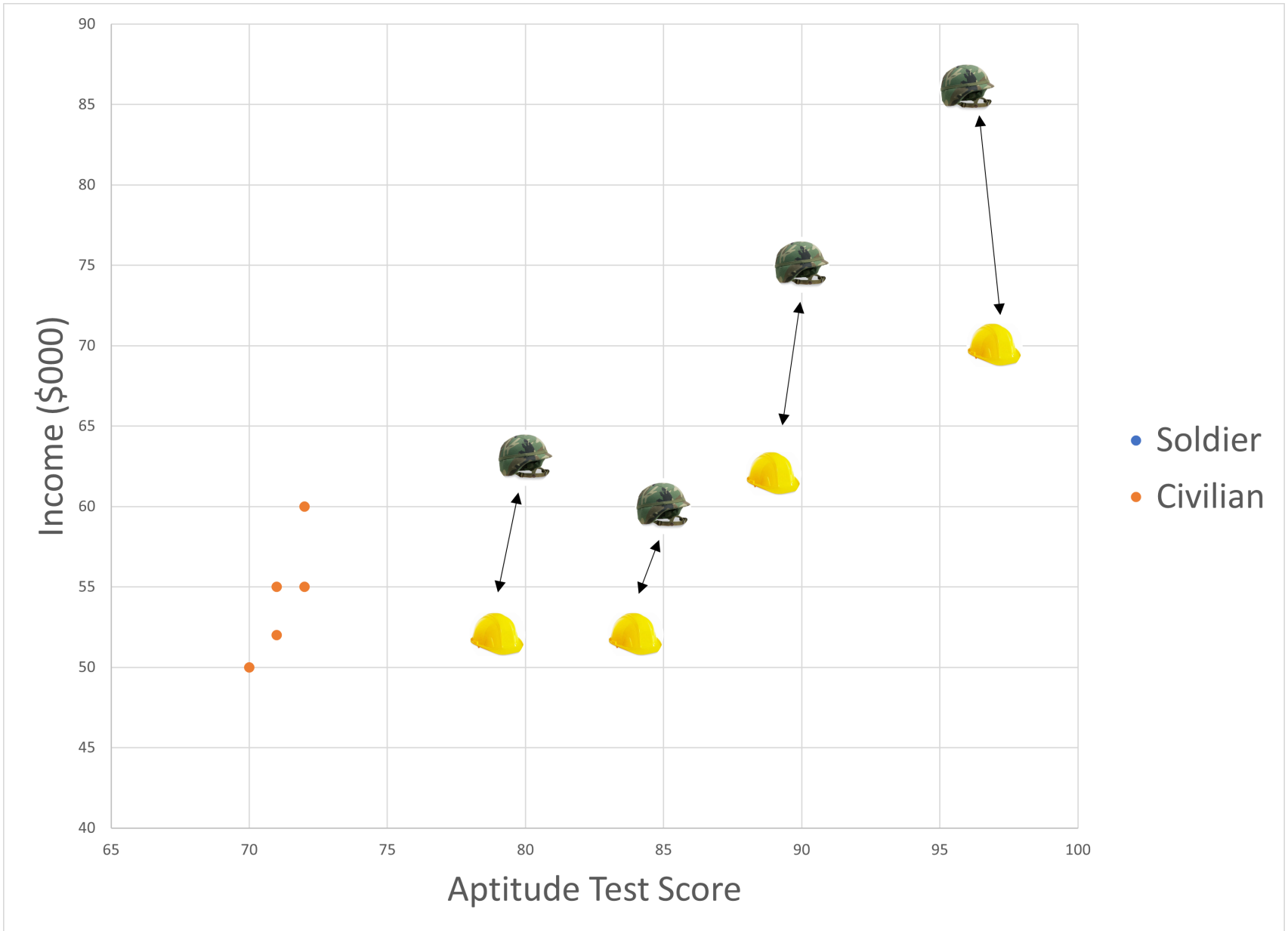
1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Nearest Neighbor Matching on Covariates

Matches are determined by similarity of Treated Covariates to Control Covariates



1-to-1 on a Single Covariate



Evaluating 1-to-1 on Single Covariate

Q: What are the (dis)advantages of 1-to-1 on Single Covariate?

Evaluating 1-to-1 on Single Covariate

Q: What are the (dis)advantages of 1-to-1 on Single Covariate?

A:

Advantages

- Intuitive
- Easy to implement. Distance measure is simply in the units of the covariate

Evaluating 1-to-1 on Single Covariate

Q: What are the (dis)advantages of 1-to-1 on Single Covariate?

A:

Advantages

- Intuitive
- Easy to implement. Distance measure is simply in the units of the covariate

Disadvantages

- Matching on a single covariate may lead to bias since that likely doesn't provide a full characterization of the outcome dynamics

k-to-1 matching on Single Covariate

Q: What the (dis)advantages of increasing k?

k-to-1 matching on Single Covariate

Q: What the (dis)advantages of increasing k?

A:

Advantage

- Improve precision by avoiding dropped observations

k-to-1 matching on Single Covariate

Q: What the (dis)advantages of increasing k?

A:

Advantage

- Improve precision by avoiding dropped observations

Disadvantage

- Can lead to poor matches. By definition, the 2nd, 3rd, etc..matches are worse than the first

k-to-1 matching on Single Covariate

Q: What the (dis)advantages of increasing k?

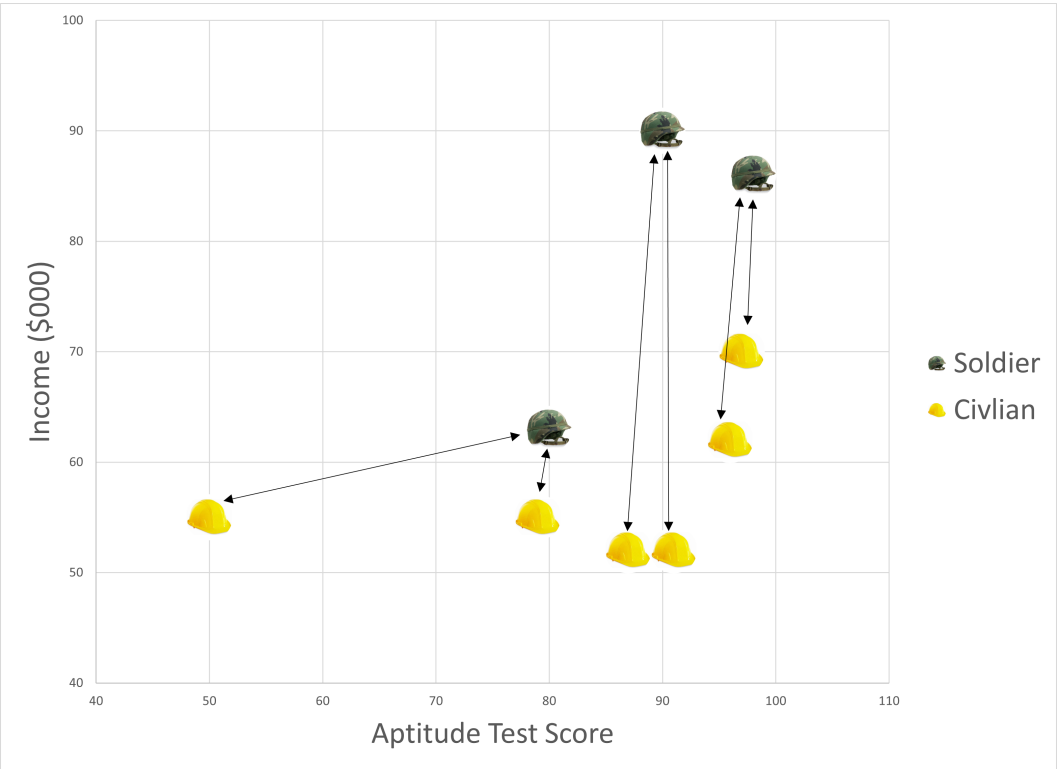
A:

Advantage

- Improve precision by avoiding dropped observations

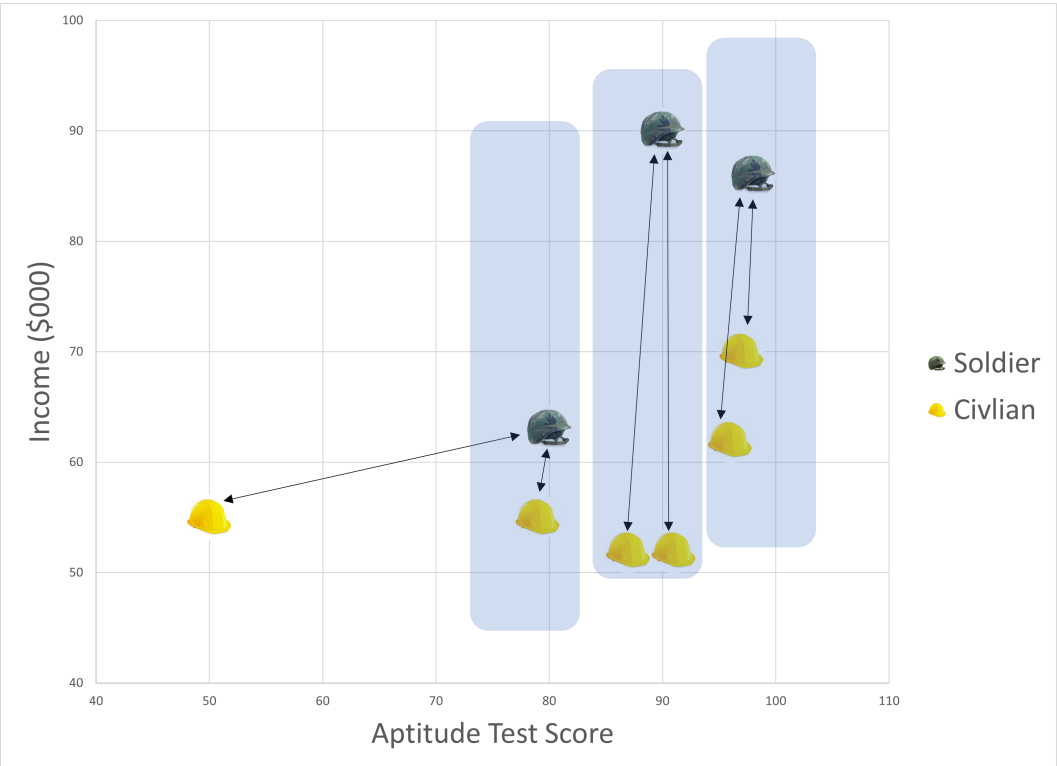
Disadvantage

- Can lead to poor matches. By definition, the 2nd, 3rd, etc..matches are worse than the first



Imposing a Caliper to Avoid Bad Matches

In our case study, the caliper excludes Civilians whose test scores are beyond a certain distance from their matched Soldier. In the illustration below, the caliper is the light blue shaded area. Notice the Civilian with the lowest test score was matched to a Soldier, but is dropped because it is outside the Caliper.



Imposing a Caliper

Q: What are the (dis)advantages of imposing a Caliper?

A:

Imposing a Caliper

Q: What are the (dis)advantages of imposing a Caliper?

A:

Advantage

- Avoid extremely poor matches

Disadvantage

- Dropped observations

Matching on multiple Covariates

Matching on multiple covariates requires a more sophisticated distance measure

Matching on multiple Covariates

Matching on multiple covariates requires a more sophisticated distance measure

$$\text{Euclidean: } \sqrt{\sum_{k=1}^{K}(X_{ki} - X_{kj})^2} = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^\intercal(\mathbf{X}_i - \mathbf{X}_j)}$$

$$\text{Normalized Euclidean: } \sqrt{\sum_{k=1}^{K}\frac{(X_{ki} - X_{kj})^2}{\hat{\sigma}_k^2}} = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^\intercal \hat{V}^{-1}(\mathbf{X}_i - \mathbf{X}_j)}$$

$$\text{Mahalanobis: } \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^\intercal \hat{\Sigma}^{-1}(\mathbf{X}_i - \mathbf{X}_j)}$$

where V is an estimate of the diagonal Variance matrix of X, Σ is an estimate of the covariance of X, and $\mathbf{X}$ is a matrix with units in the rows and different covariates in the columns

Example of Computing a Distance Metric

Compute Norm.Euclidean distance of each control to find k=2 matches for Soldier ID=1

ID	D(Soldier=1)	Test Score	Yrs School Completed	Income
1	1	83	12	63
2	1	70	12	70
3	1	89	16	65
4	0	71	13	55
5	0	69	12	60
6	0	81	12	62
7	0	83	11	48
8	0	89	14	55
9	0	89	14	47
	Sigma	8.38	1.54	

Example of Computing a Distance Metric

Compute Norm.Euclidean distance of each control to find k=2 matches for Soldier ID=1

ID	D(Soldier=1)	Test Score	Yrs School Completed	Income
1	1	83	12	63
2	1	70	12	70
3	1	89	16	65
4	0	71	13	55
5	0	69	12	60
6	0	81	12	62
7	0	83	11	48
8	0	89	14	55
9	0	89	14	47
	Sigma	8.38	1.54	

Normalized Euclidean Distance Relative to ID=1 for k=2

	Test Score			Yrs School Completed			
ID	Difference	Squared	Normalized	Difference	Squared	Normalized	Sum
4	12	144	17.18	-1	1	0.65	17.83
5	14	196	23.38	0	0	0.00	23.38
6	2	4	0.48	0	0	0.00	0.48
7	0	0	0.00	1	1	0.65	0.65
8	-6	36	4.29	-2	4	2.60	6.90
9	-6	36	4.29	-2	4	2.60	6.90

Example of Computing a Distance Metric

Compute Norm.Euclidean distance of each control to find k=2 matches for Soldier ID=1

ID	D(Soldier=1)	Test Score	Yrs School Completed	Income
1	1	83	12	63
2	1	70	12	70
3	1	89	16	65
4	0	71	13	55
5	0	69	12	60
6	0	81	12	62
7	0	83	11	48
8	0	89	14	55
9	0	89	14	47
	Sigma	8.38	1.54	

Normalized Euclidean Distance Relative to ID=1 for k=2

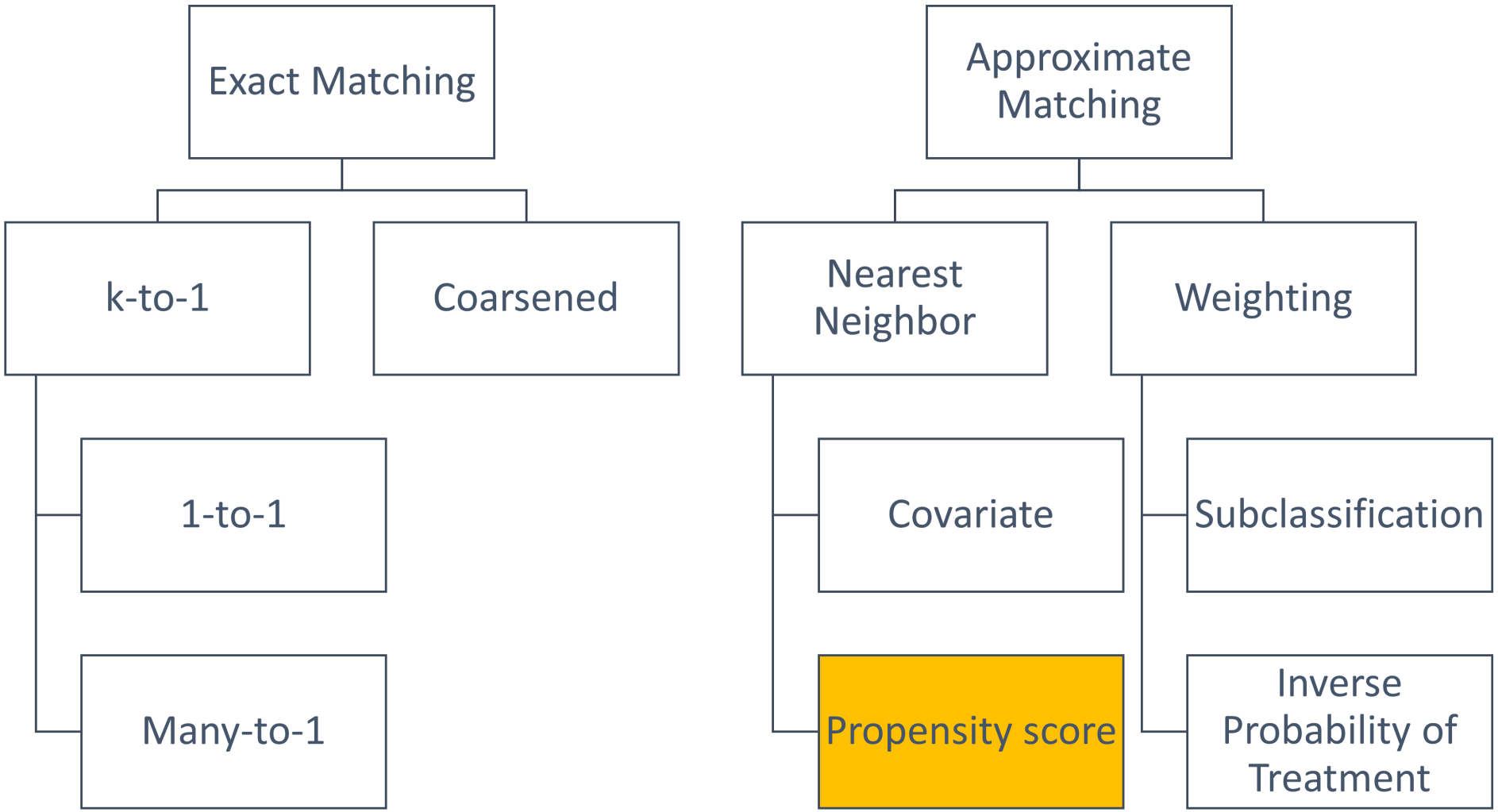
	Test Score			Yrs School Completed			
ID	Difference	Squared	Normalized	Difference	Squared	Normalized	Sum
4	12	144	17.18	-1	1	0.65	17.83
5	14	196	23.38	0	0	0.00	23.38
6	2	4	0.48	0	0	0.00	0.48
7	0	0	0.00	1	1	0.65	0.65
8	-6	36	4.29	-2	4	2.60	6.90
9	-6	36	4.29	-2	4	2.60	6.90

- Control units ID=6,7 are suitable matches for Soldier ID=1 since these have the smallest normalized Euclidean distance
- The variances here were computed using the full sample, which is suitable for ATE. If your focus is ATT, then use only the control sample. (See Stuart 2010 Section 2.2)

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Nearest Neighbor Matching via Propensity Scores

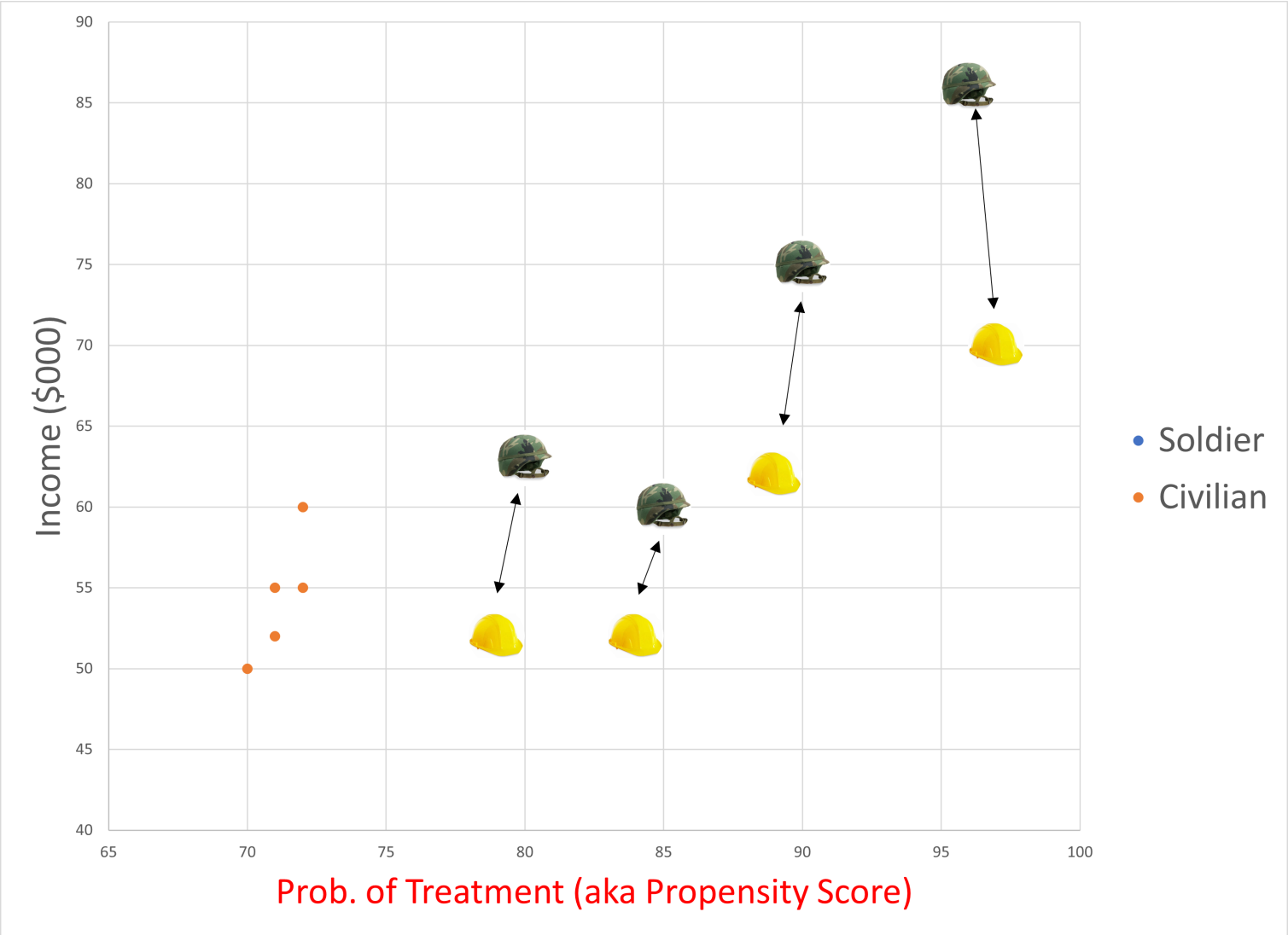


Matching on Propensity Scores

Match on the probability of treatment (aka Propensity Score)

Matching on Propensity Scores

Match on the probability of treatment (aka Propensity Score)



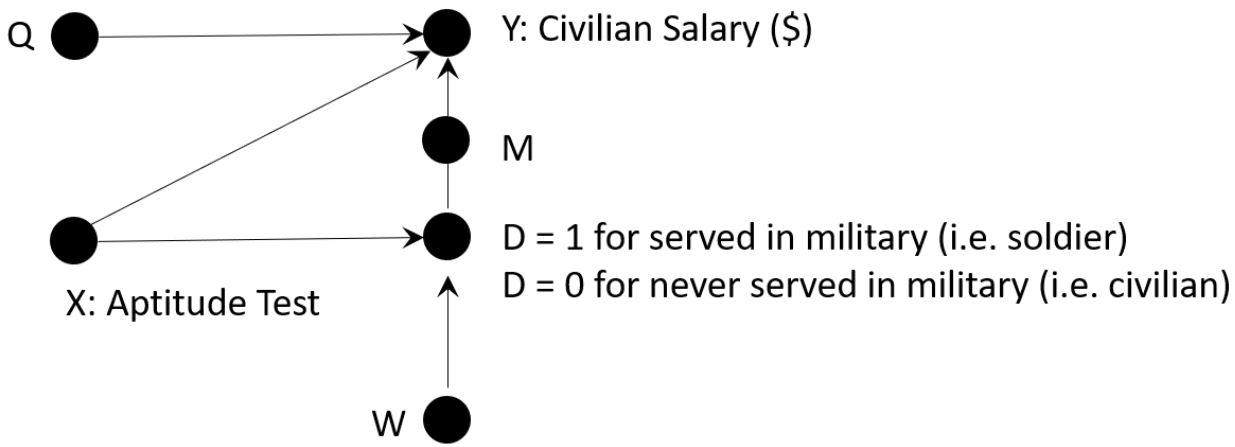
Estimating Propensity scores

- Propensity Scores are estimated via binary outcome models (e.g. logit, probit, neural networks with sigmoid activation, etc..)
- For instance, using the logit link function yields:
  $Pr[D_i = 1|X_i] = \frac{e^{\mathbf{x}_i^\intercal \beta}}{1+e^{\mathbf{x}_i^\intercal \beta}}$

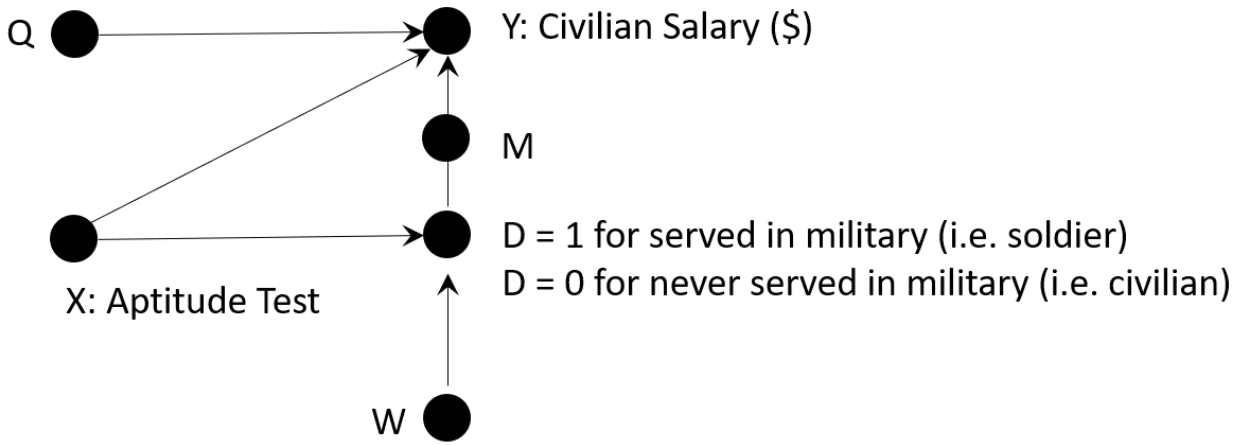
Estimating Propensity scores

- Propensity Scores are estimated via binary outcome models (e.g. logit, probit, neural networks with sigmoid activation, etc..)
- For instance, using the logit link function yields:
  $Pr[D_i = 1|X_i] = \frac{e^{\mathbf{x}_i^\intercal \beta}}{1+e^{\mathbf{x}_i^\intercal \beta}}$
- Notice that this is NOT a model of the outcome variable Y, but rather a model of treatment D
- We typically use the pre-treatment covariates X to estimate the probability of treatment since this is the information that would be available for a prediction of treatment for new units

Choosing Appropriate Covariates for Propensity Scores

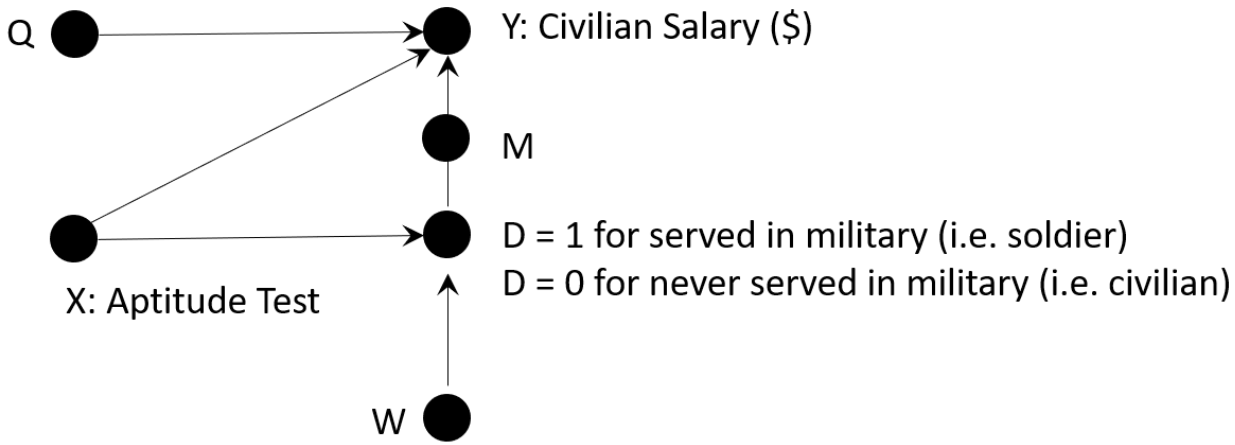


Choosing Appropriate Covariates for Propensity Scores



Q: Which variables should we include in the propensity score model?

Choosing Appropriate Covariates for Propensity Scores



Q: Which variables should we include in the propensity score model?
A:

- Include i) "Good Controls" (X), ii) drivers of the outcome variable (Q) which reduces noise in estimate of Y
- Avoid i) drivers of treatment that are not related to outcome (W) since doing so would lower precision Smith and Todd 2005, ii) mediators (M) since doing so may attenuate the treatment effect estimate

Assumptions for Propensity Score Matching

- Overlap Assumption $0 < Pr[D = 1|X] < 1 \rightarrow$ For each each value of X there are treated and control units.

Assumptions for Propensity Score Matching

- Overlap Assumption $0 < Pr[D = 1|X] < 1 \rightarrow$ For each each value of X there are treated and control units.
- $\{Y_0, Y_1\} \perp D|p(X) \rightarrow$ Treatment assignment is as good as random once condition on the propensity score

Matching with Propensity Scores

- Use simple distance measure applied to propensity score to find the nearest k controls to each treated

Matching with Propensity Scores

- Use simple distance measure applied to propensity score to find the nearest k controls to each treated
- Similar to covariate matching
  - a larger k may increase precision by increasing the number of matches, but also may increase bias by including worse matches
  - controls can be matched with or without replacement
  - can impose a caliper to avoid poor bad matches
  - Treatment Effects are estimated via the average of the CATEs

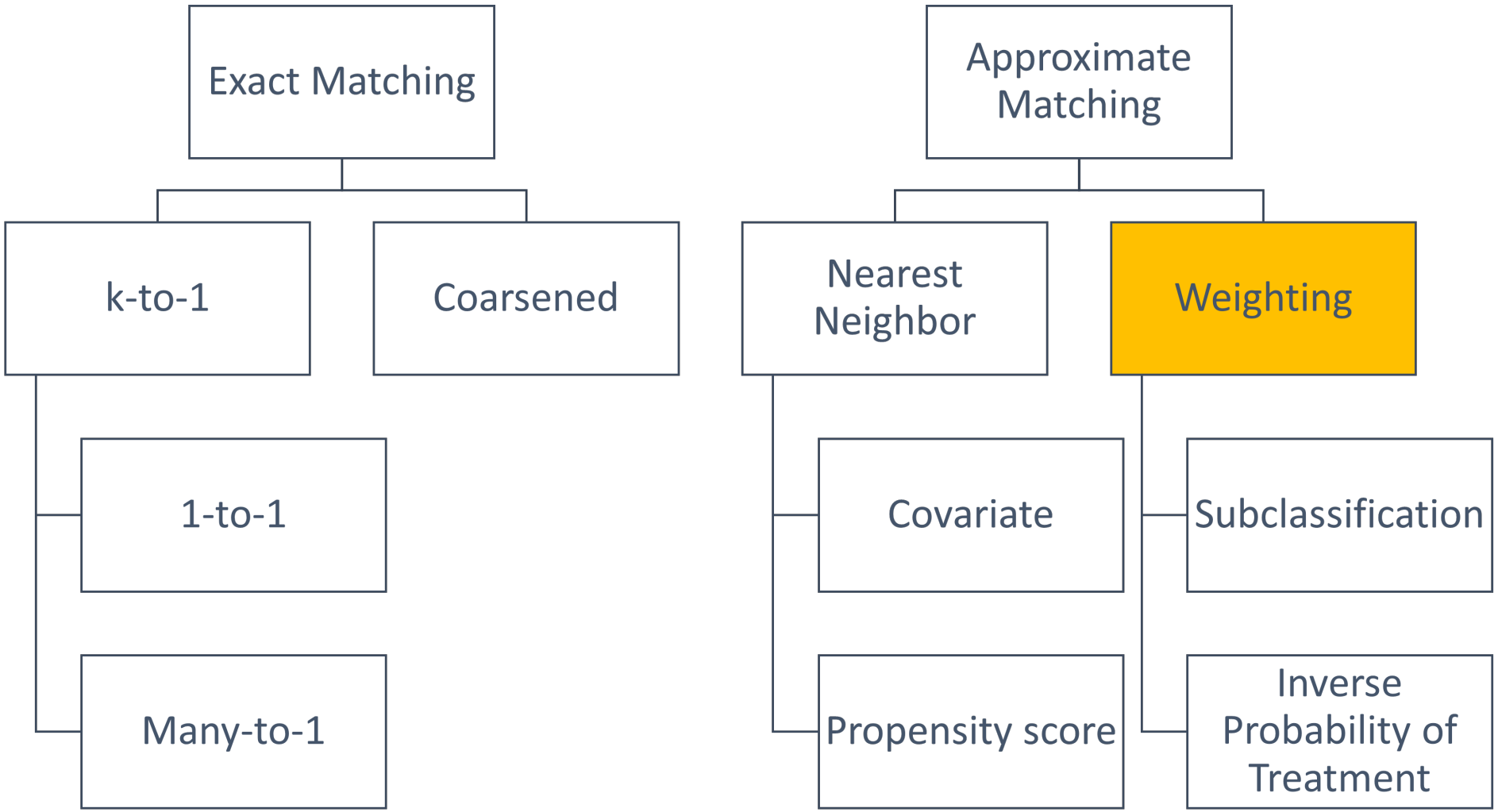
Matching with Propensity Scores

- Use simple distance measure applied to propensity score to find the nearest k controls to each treated
- Similar to covariate matching
  - a larger k may increase precision by increasing the number of matches, but also may increase bias by including worse matches
  - controls can be matched with or without replacement
  - can impose a caliper to avoid poor bad matches
  - Treatment Effects are estimated via the average of the CATEs
- We still need a proper set of covariates, whether matching via covariates directly or via propensity scores

Outline

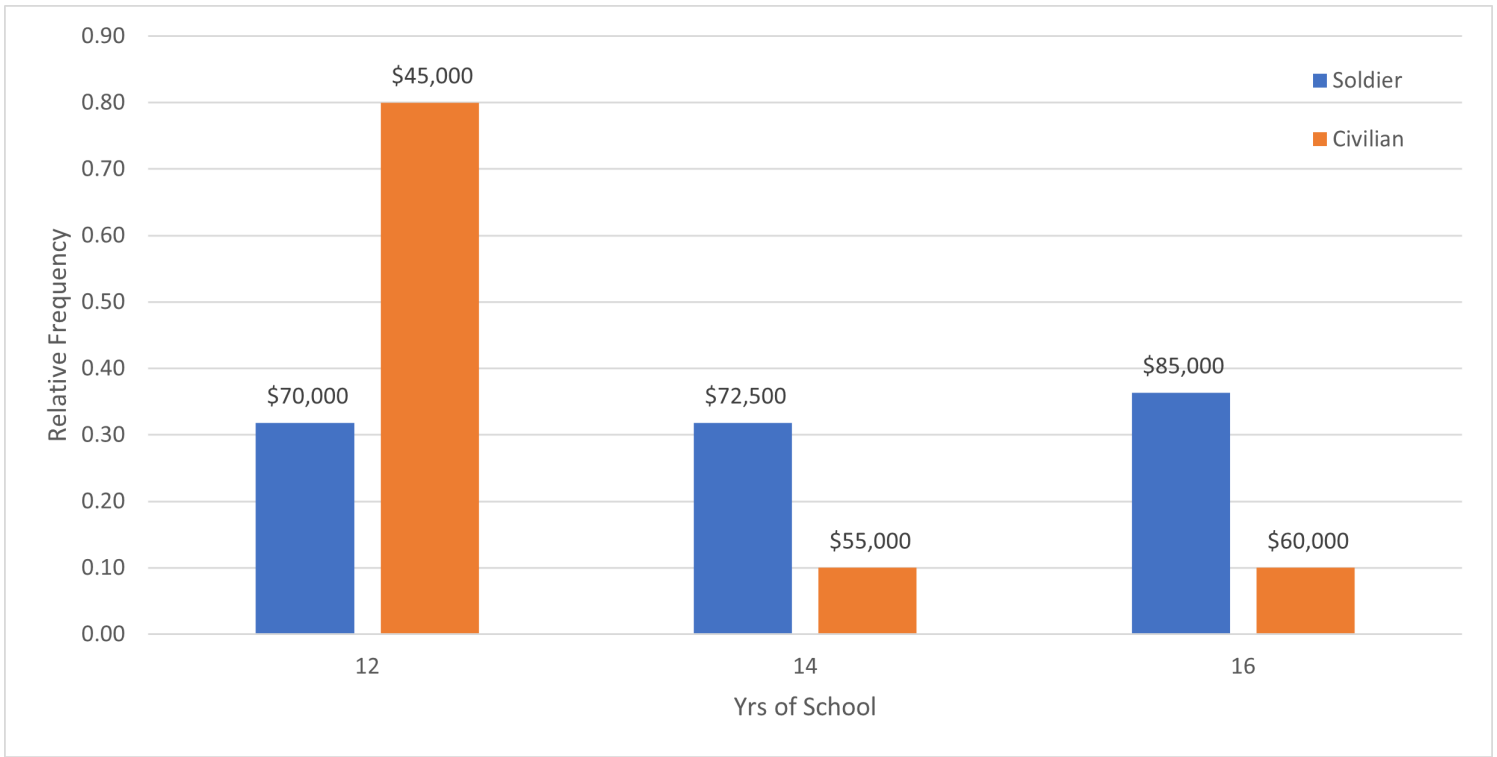
1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Weighting


Exact Matching
k-to-1
Coarsened
1-to-1
Many-to-1
Approximate Matching
Nearest Neighbor
Weighting
Covariate
Propensity score
Subclassification
Inverse Probability of Treatment

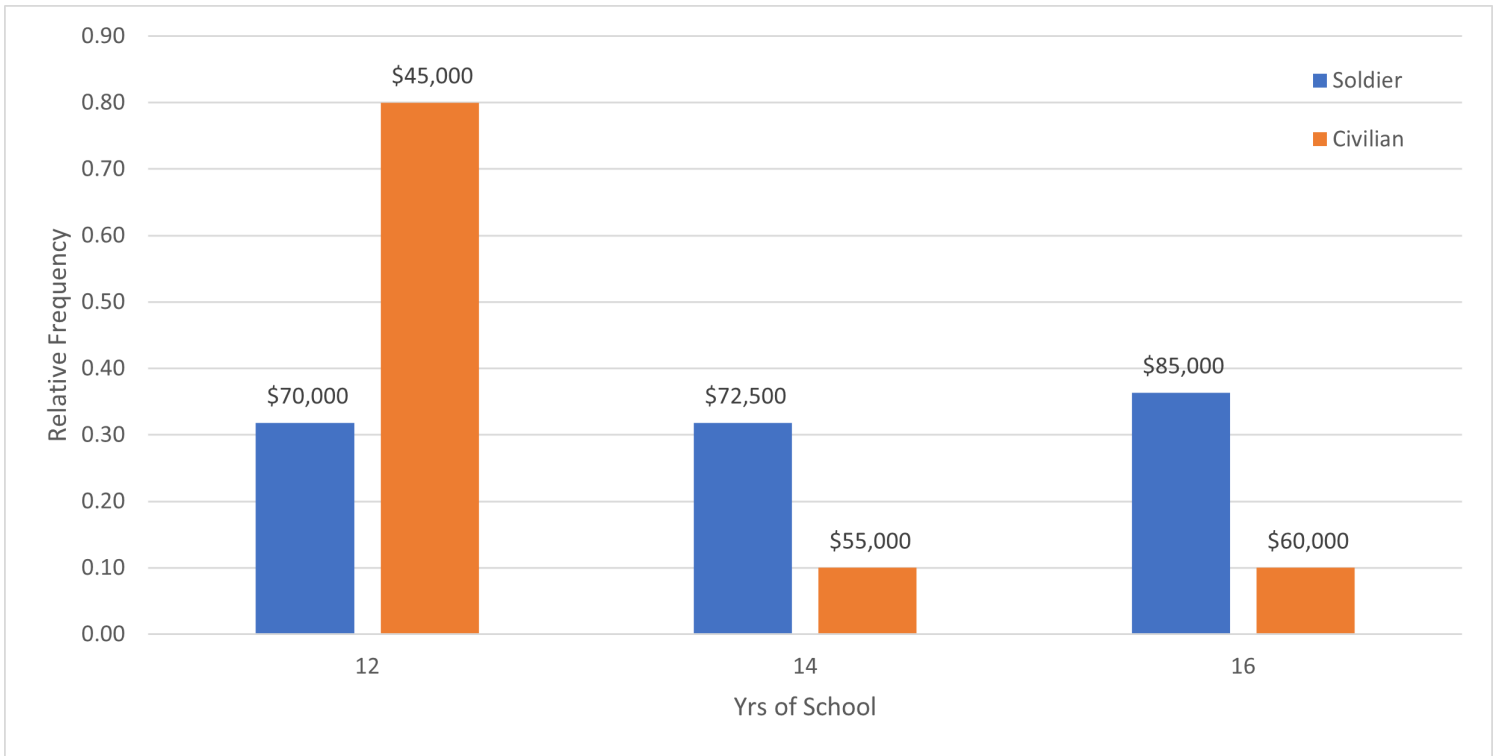
Covariate Imbalance

Q: How does the histogram illustrate potential for covariate imbalance and selection bias?



Covariate Imbalance

Q: How does the histogram illustrate potential for covariate imbalance and selection bias?



A: The DiM TE estimate $\bar{Y}_{Soldier} - \bar{Y}_{Civlian}$ might capture the effect of treatment, but also reflects the fact that most civilians only completed the 12th grade, whereas many soldiers went to college.

Weighting for Causal Effects

- The weighting approach attempts to overcome the covariate imbalance by altering the importance of each observation

Weighting for Causal Effects

- The weighting approach attempts to overcome the covariate imbalance by altering the importance of each observation
- Shortcoming of simple DiM (Difference in Means): gives all units equal importance

Weighting for Causal Effects

- The weighting approach attempts to overcome the covariate imbalance by altering the importance of each observation
- Shortcoming of simple DiM (Difference in Means): gives all units equal importance
- Shortcomings of Nearest Neighbor Matching: don't use all the observations

Weighting for Causal Effects

- The weighting approach attempts to overcome the covariate imbalance by altering the importance of each observation
- Shortcoming of simple DiM (Difference in Means): gives all units equal importance
- Shortcomings of Nearest Neighbor Matching: don't use all the observations
- Weighting methods attempt to use all the observations by explicitly providing importance weights to each observation

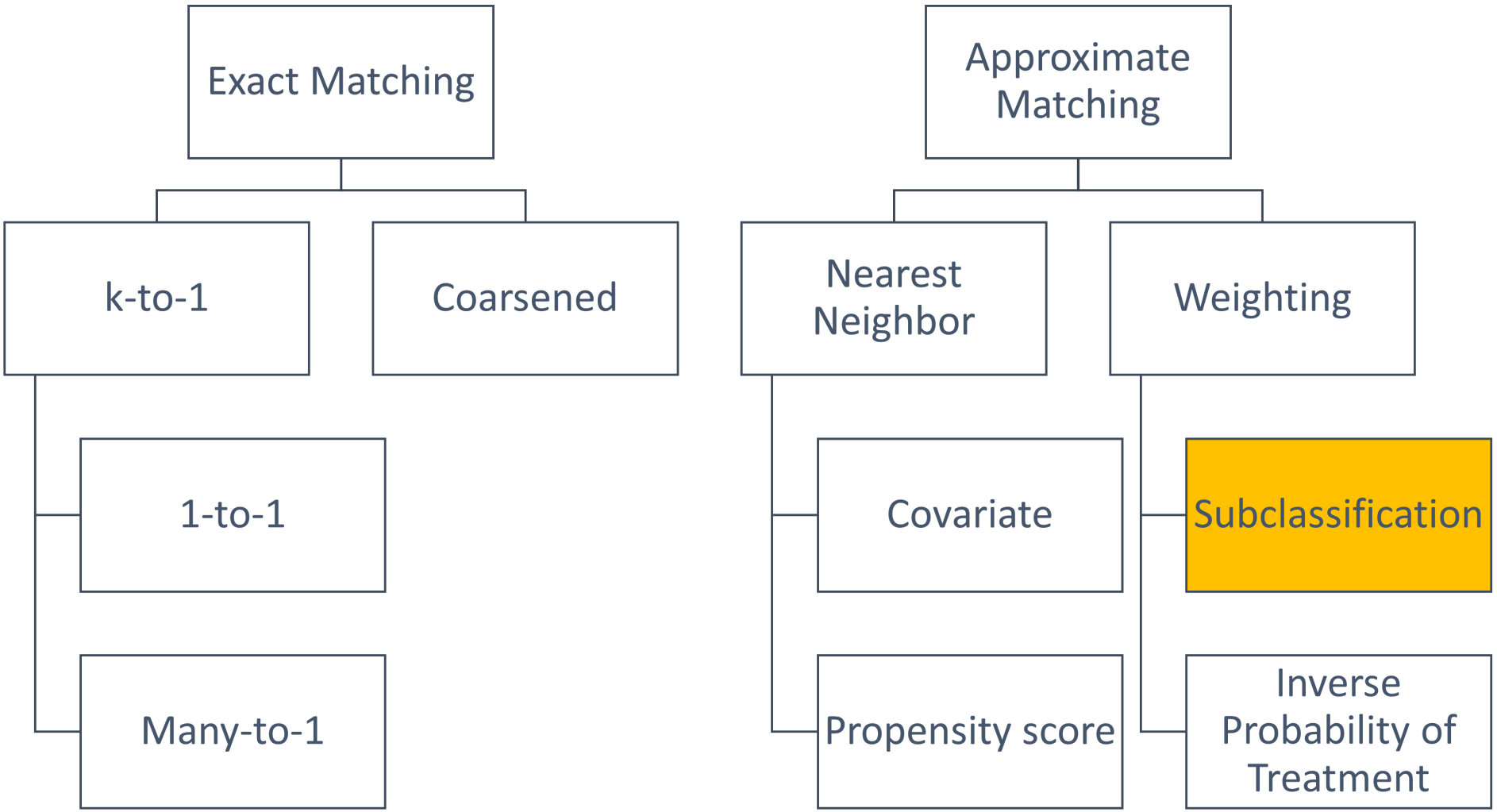
Weighting for Causal Effects

- The weighting approach attempts to overcome the covariate imbalance by altering the importance of each observation
- Shortcoming of simple DiM (Difference in Means): gives all units equal importance
- Shortcomings of Nearest Neighbor Matching: don't use all the observations
- Weighting methods attempt to use all the observations by explicitly providing importance weights to each observation
- Weighting is not "Matching" per se. However, it relies on similar tools, and thus is typically associated with Matching methods for causal effects.

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

Subclassification



Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling
   - This can be extended to multiple covariates or propensity scores
   - Full Matching is a variant that produces strata automatically

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling
   - This can be extended to multiple covariates or propensity scores
   - Full Matching is a variant that produces strata automatically
2. Compute the densities of Soldiers across years of school. Repeat for Civilians. Repeat for the pooled treat/control group.

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling
   - This can be extended to multiple covariates or propensity scores
   - Full Matching is a variant that produces strata automatically
2. Compute the densities of Soldiers across years of school. Repeat for Civilians. Repeat for the pooled treat/control group.
3. For ATT: Compute the weighted average income of Soldiers using the Soldier's density. Compute the weighted average income of Civilians using the Soldier's density. Subtract the Civilian weighted average income from the Soldier's.

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling
   - This can be extended to multiple covariates or propensity scores
   - Full Matching is a variant that produces strata automatically
2. Compute the densities of Soldiers across years of school. Repeat for Civilians. Repeat for the pooled treat/control group.
3. For ATT: Compute the weighted average income of Soldiers using the Soldier's density. Compute the weighted average income of Civilians using the Soldier's density. Subtract the Civilian weighted average income from the Soldier's.
4. For ATC: Compute the weighted average income of Soldiers using the Civilian's density. Compute the weighted average income of Civilians using the Civilian's density. Subtract the Civilian weighted average income from the Soldier's.

Subclassification (aka Stratification) Steps

Overcome selection bias by assuming Civilians' schooling was similar to soldiers

1. Stratify the data by yrs of schooling
   - This can be extended to multiple covariates or propensity scores
   - Full Matching is a variant that produces strata automatically
2. Compute the densities of Soldiers across years of school. Repeat for Civilians. Repeat for the pooled treat/control group.
3. For ATT: Compute the weighted average income of Soldiers using the Soldier's density. Compute the weighted average income of Civilians using the Soldier's density. Subtract the Civilian weighted average income from the Soldier's.
4. For ATC: Compute the weighted average income of Soldiers using the Civilian's density. Compute the weighted average income of Civilians using the Civilian's density. Subtract the Civilian weighted average income from the Soldier's.
5. For ATE: $ATT \times \frac{N_{Treated}}{N_{Pooled}} + ATC \times \frac{N_{Control}}{N_{Pooled}}$

Subclassification Example - Frequency of Soldier/Civilian

Q: Can you interpret ATT and it's components?

				Density			Income			
Yrs of Schooling	Soldier	Civilian	Sum	Soldier	Civilian	Pooled	Soldier	Civilian		
12	35	160	195	0.32	0.80	0.63	$ 70,000	$ 45,000		
14	35	20	55	0.32	0.10	0.18	$ 72,500	$ 55,000		
16	40	20	60	0.36	0.10	0.19	$ 85,000	$ 60,000		
Total	110	200	310	1	1	1				
Rel. Frequency	0.35	0.65	1							
				Unconditional Avg			$ 75,833	$ 53,333	DiM:	$ 22,500
							Adjusted Values			
				Weight by Soldier Density			$ 76,250	$ 53,636	ATT:	$ 22,614
				Weight by Civilian Density			$ 71,750	$ 47,500	ATC:	$ 24,250
									ATE:	$ 23,669

Subclassification Example - Frequency of Soldier/Civilian

Q: Can you interpret ATT and it's components?

				Density			Income			
Yrs of Schooling	Soldier	Civilian	Sum	Soldier	Civilian	Pooled	Soldier	Civilian		
12	35	160	195	0.32	0.80	0.63	$ 70,000	$ 45,000		
14	35	20	55	0.32	0.10	0.18	$ 72,500	$ 55,000		
16	40	20	60	0.36	0.10	0.19	$ 85,000	$ 60,000		
Total	110	200	310	1	1	1				
Rel. Frequency	0.35	0.65	1							
				Unconditional Avg			$ 75,833	$ 53,333	DiM:	$ 22,500
							Adjusted Values			
				Weight by Soldier Density			$ 76,250	$ 53,636	ATT:	$ 22,614
				Weight by Civilian Density			$ 71,750	$ 47,500	ATC:	$ 24,250
									ATE:	$ 23,669

A:

- The weighted average Soldier income is $\$76,250$

Subclassification Example - Frequency of Soldier/Civilian

Q: Can you interpret ATT and it's components?

				Density			Income			
Yrs of Schooling	Soldier	Civilian	Sum	Soldier	Civilian	Pooled	Soldier	Civilian		
12	35	160	195	0.32	0.80	0.63	$ 70,000	$ 45,000		
14	35	20	55	0.32	0.10	0.18	$ 72,500	$ 55,000		
16	40	20	60	0.36	0.10	0.19	$ 85,000	$ 60,000		
Total	110	200	310	1	1	1				
Rel. Frequency	0.35	0.65	1							
				Unconditional Avg			$ 75,833	$ 53,333	DiM:	$ 22,500
							Adjusted Values			
				Weight by Soldier Density			$ 76,250	$ 53,636	ATT:	$ 22,614
				Weight by Civilian Density			$ 71,750	$ 47,500	ATC:	$ 24,250
									ATE:	$ 23,669

A:

- The weighted average Soldier income is $\$76,250$
- If Civilians had same population distribution has Soldiers, their average income would be $\$53,636$

Subclassification Example - Frequency of Soldier/Civilian

Q: Can you interpret ATT and it's components?

				Density			Income			
Yrs of Schooling	Soldier	Civilian	Sum	Soldier	Civilian	Pooled	Soldier	Civilian		
12	35	160	195	0.32	0.80	0.63	$ 70,000	$ 45,000		
14	35	20	55	0.32	0.10	0.18	$ 72,500	$ 55,000		
16	40	20	60	0.36	0.10	0.19	$ 85,000	$ 60,000		
Total	110	200	310	1	1	1				
Rel. Frequency	0.35	0.65	1							
				Unconditional Avg			$ 75,833	$ 53,333	DiM:	$ 22,500
							Adjusted Values			
				Weight by Soldier Density			$ 76,250	$ 53,636	ATT:	$ 22,614
				Weight by Civilian Density			$ 71,750	$ 47,500	ATC:	$ 24,250
									ATE:	$ 23,669

A:

- The weighted average Soldier income is $\$76,250$
- If Civilians had same population distribution has Soldiers, their average income would be $\$53,636$
- ATT is the $\$22,614$ difference in these weighted average incomes

Subclassification Example - Frequency of Soldier/Civilian

Q: Can you interpret ATT and it's components?

				Density			Income			
Yrs of Schooling	Soldier	Civilian	Sum	Soldier	Civilian	Pooled	Soldier	Civilian		
12	35	160	195	0.32	0.80	0.63	$ 70,000	$ 45,000		
14	35	20	55	0.32	0.10	0.18	$ 72,500	$ 55,000		
16	40	20	60	0.36	0.10	0.19	$ 85,000	$ 60,000		
Total	110	200	310	1	1	1				
Rel. Frequency	0.35	0.65	1							
				Unconditional Avg			$ 75,833	$ 53,333	DiM:	$ 22,500
							Adjusted Values			
				Weight by Soldier Density			$ 76,250	$ 53,636	ATT:	$ 22,614
				Weight by Civilian Density			$ 71,750	$ 47,500	ATC:	$ 24,250
									ATE:	$ 23,669

A:

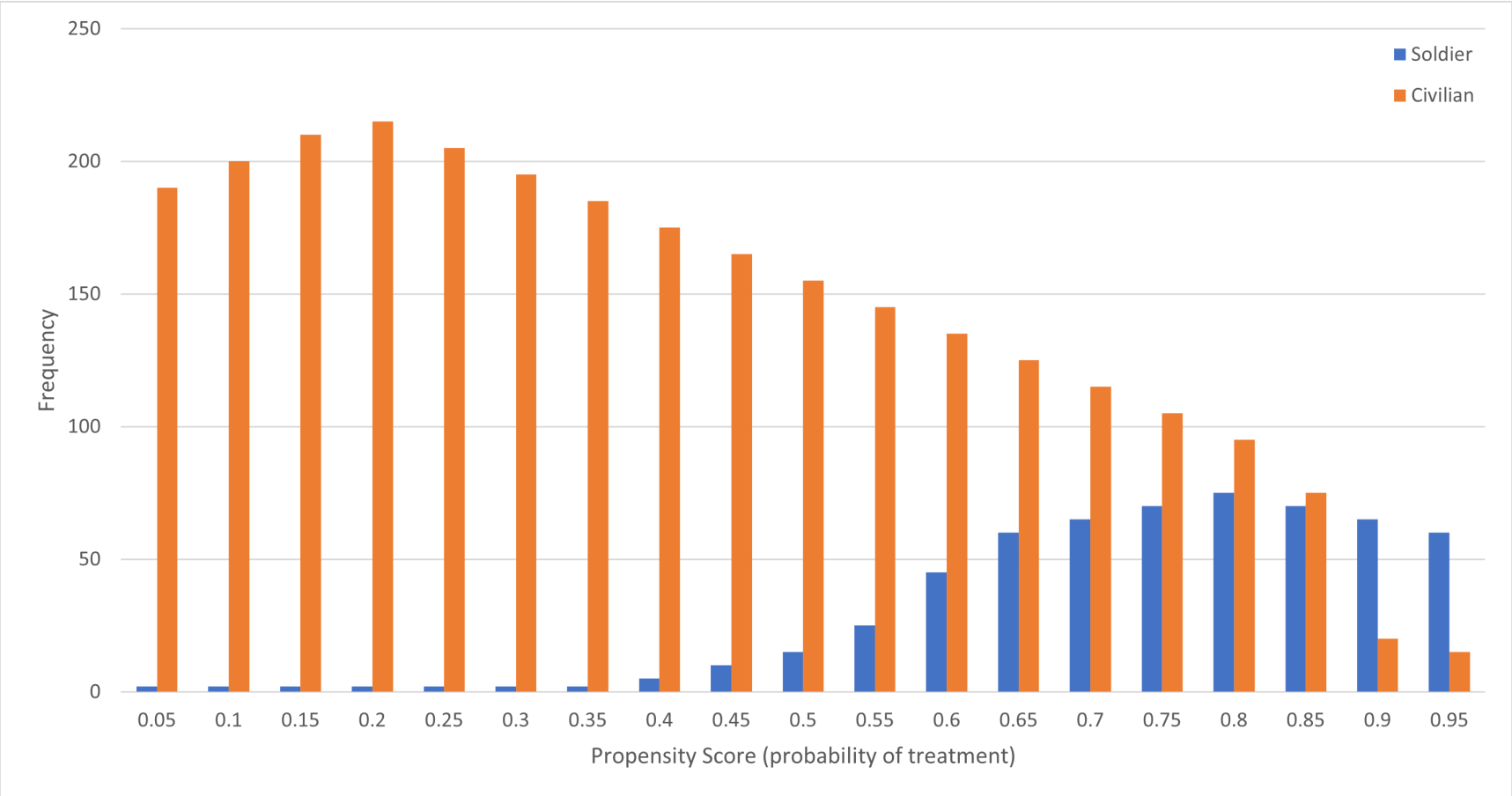
- The weighted average Soldier income is $\$76,250$
- If Civilians had same population distribution has Soldiers, their average income would be $\$53,636$
- ATT is the $\$22,614$ difference in these weighted average incomes
- Implying that being a Soldier enhanced earnings by $\$22,614$ relative to a Civilian with similar years of schooling

Outline

1. Case Study - Motivation
2. Matching
3. Exact Matching
   - 1-to-1
   - many-to-1
   - Coarsened Exact Matching
   - (Dis)advantages of Exact Matching
4. Approximate Matching - Nearest Neighbor
   - Covariate Matching
   - Propensity Score Matching
5. Approximate Matching - Weighting
   - Subclassification
   - Inverse Probability of Treatment (IPTW)

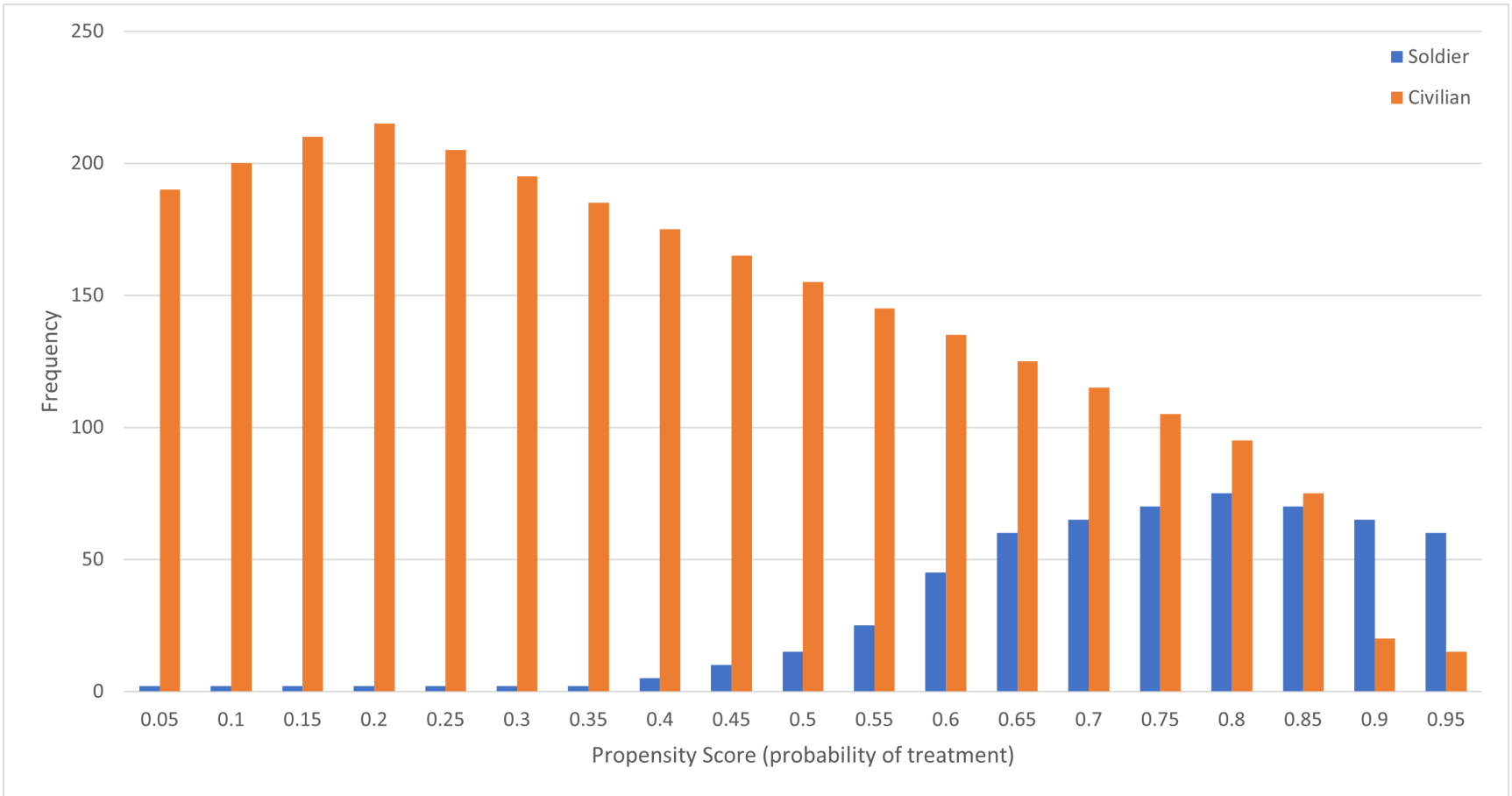
Propensity Density

Q: Who are the "rare" Soldiers?



Propensity Density

Q: Who are the "rare" Soldiers?



A: Soldiers with low estimated propensity scores are rare. Their existence can strengthen our model.

IPTW Intuition

IPTW Intuition

- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?

IPTW Intuition

- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?
- A: The DiM estimate will be skewed toward the TE of the soldiers with higher propensity scores, since there are more of them in the sample. This may generate selection bias because

IPTW Intuition

- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?
- A: The DiM estimate will be skewed toward the TE of the soldiers with higher propensity scores, since there are more of them in the sample. This may generate selection bias because these soldiers may have different characteristics than the general population of soldiers, making our model unsuitable for all of the treated units.

IPTW Intuition

- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?
- A: The DiM estimate will be skewed toward the TE of the soldiers with higher propensity scores, since there are more of them in the sample. This may generate selection bias because these soldiers may have different characteristics than the general population of soldiers, making our model unsuitable for all of the treated units.
- Q: How does Inverse Probability of Treatment Weighting (IPTW) resolve this problem?

IPTW Intuition

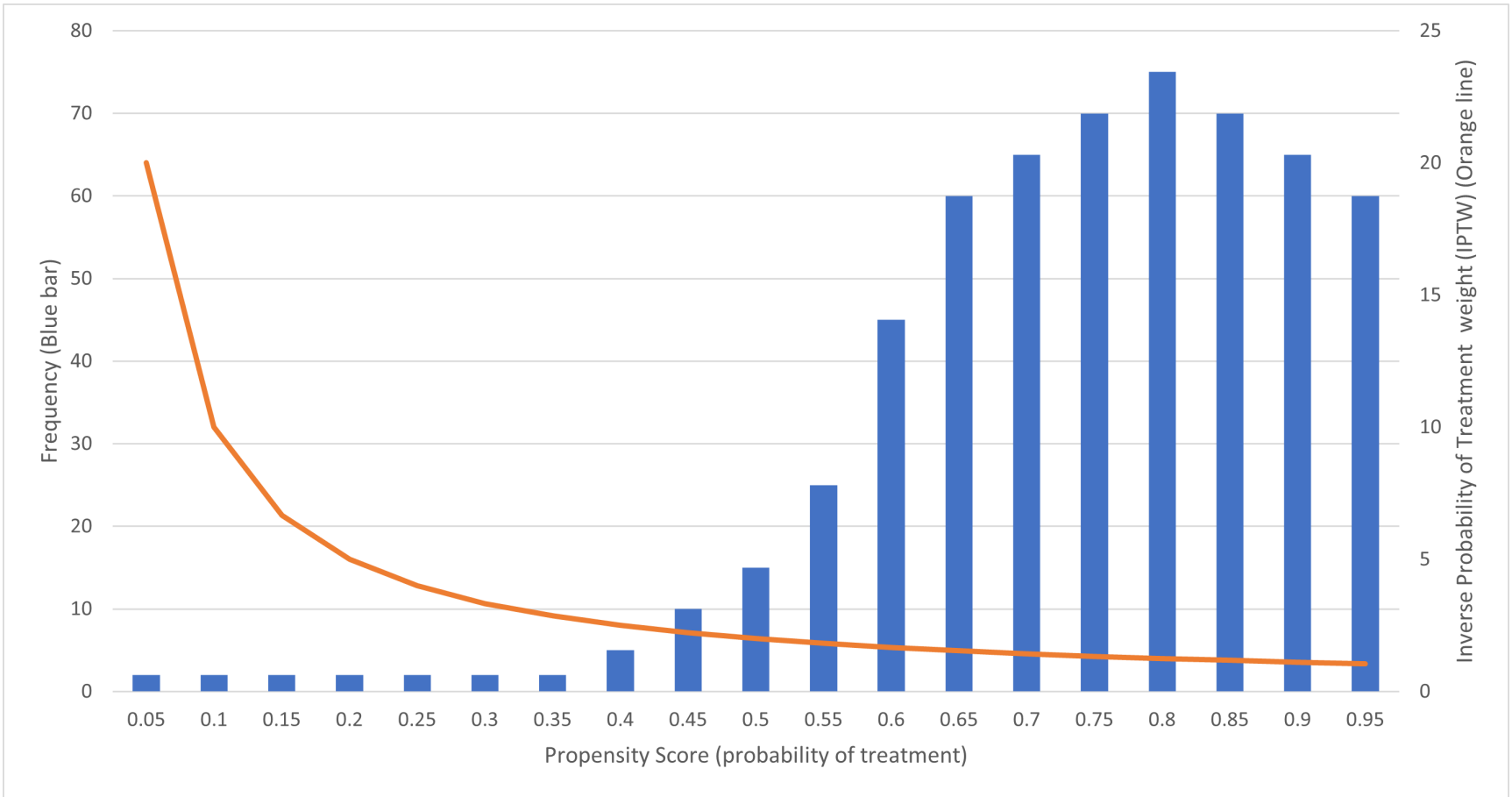
- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?
- A: The DiM estimate will be skewed toward the TE of the soldiers with higher propensity scores, since there are more of them in the sample. This may generate selection bias because these soldiers may have different characteristics than the general population of soldiers, making our model unsuitable for all of the treated units.
- Q: How does Inverse Probability of Treatment Weighting (IPTW) resolve this problem?
- A: IPTW reduces the weight of Soldiers with high propensity score and increases the weight of Soldiers with low propensity scores. In doing so, IPTW increases the per unit importance of the rare observations and lowers the per unit importance of the common observations.

IPTW Intuition

- Q: Given the density of Soldier propensity scores, how might a simple DiM TE estimate generate a selection bias?
- A: The DiM estimate will be skewed toward the TE of the soldiers with higher propensity scores, since there are more of them in the sample. This may generate selection bias because these soldiers may have different characteristics than the general population of soldiers, making our model unsuitable for all of the treated units.
- Q: How does Inverse Probability of Treatment Weighting (IPTW) resolve this problem?
- A: IPTW reduces the weight of Soldiers with high propensity score and increases the weight of Soldiers with low propensity scores. In doing so, IPTW increases the per unit importance of the rare observations and lowers the per unit importance of the common observations.
- A similar logic and process is applied to the Civilians.

IPTW Weights

Frequency and IPTW Weights for Solders



IPTW augments DiM

Recall a simple Difference In Means (DiM) estimate of ATE:

$$DiM = \frac{1}{N}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}(1 - D_i) Y_i$$

IPTW augments DiM

Recall a simple Difference In Means (DiM) estimate of ATE:

$$DiM = \frac{1}{N}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}(1 - D_i) Y_i$$

IPTW imposes a weighting

$$IPTW = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right) D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1 - D_i) Y_i$$

where $\hat{e}_i$ is an estimate of the propensity score for unit i.

Refining the IPTW estimate of ATE

$$IPTW = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right) D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1-D_i)Y_i$$

Refining the IPTW estimate of ATE

$$IPTW = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right) D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1-D_i)Y_i$$

In practice, this IPTW estimate is shown to have large variance. Hirano, Imbens, and Ridder (2003) suggest a slight adjustment

$$IPTW = \left(\sum_{i=1}^{N}\frac{D_i}{\hat{e}_i}\right)^{-1}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right) D_i Y_i - \left(\sum_{i=1}^{N}\frac{1-D_i}{1-\hat{e}_i}\right)^{-1}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1-D_i)Y_i$$

IPTW Edge Cases

- Q: Suppose the propensity scores are near 0 or 1. Why is this cited as a disadvantage of the IPTW estimator?

IPTW Edge Cases

- Q: Suppose the propensity scores are near 0 or 1. Why is this cited as a disadvantage of the IPTW estimator?
- A: Consider a Soldier with propensity score near 0. Their weight becomes very large.

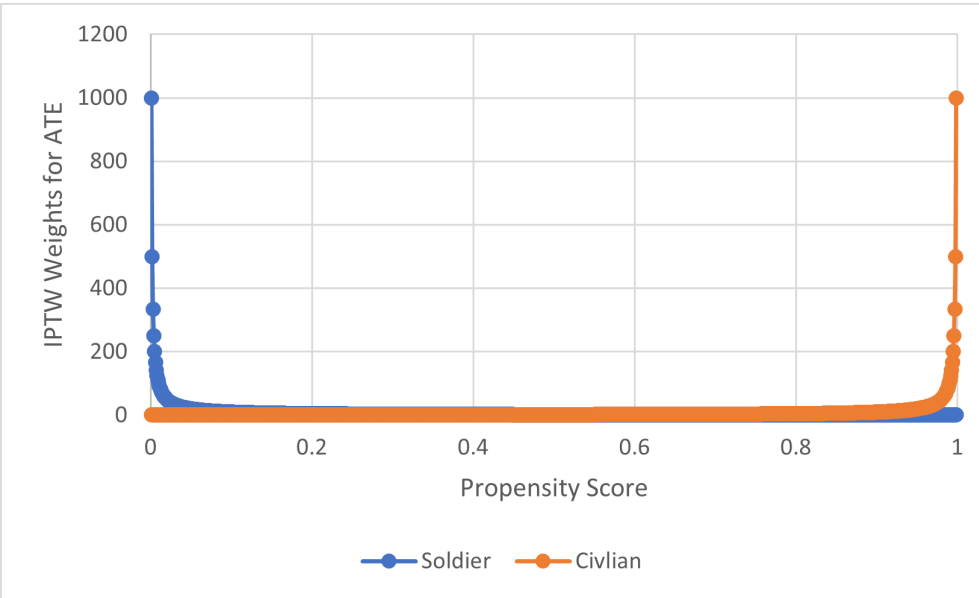
IPTW Edge Cases

- Q: Suppose the propensity scores are near 0 or 1. Why is this cited as a disadvantage of the IPTW estimator?
- A: Consider a Soldier with propensity score near 0. Their weight becomes very large. Likewise, the weight of a Civilian with a propensity score near 1 carries a very large weight.

IPTW Edge Cases

- Q: Suppose the propensity scores are near 0 or 1. Why is this cited as a disadvantage of the IPTW estimator?
- A: Consider a Soldier with propensity score near 0. Their weight becomes very large. Likewise, the weight of a Civilian with a propensity score near 1 carries a very large weight.
- One approach to stabilize the weighting scheme is by truncating the weights (e.g. at the 1^{st} and 99^{th} quantiles of the weight distribution)

Resolving Edge Cases with Truncating





IPTW for ATT

Recall

$$IPTW = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right) D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1-D_i)Y_i$$

To derive an estimator for ATT it is useful to see the ATE weights as

$$w_{ATE} = \frac{D_i}{\hat{e}_i} + \frac{1-D_i}{1-\hat{e}_i}$$

IPTW for ATT

Recall

$$IPTW = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\hat{e}_i}\right)D_i Y_i - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1-\hat{e}_i}\right)(1-D_i)Y_i$$

To derive an estimator for ATT it is useful to see the ATE weights as

$$w_{ATE} = \frac{D_i}{\hat{e}_i} + \frac{1-D_i}{1-\hat{e}_i}$$

The ATT weights are given as

$$w_{ATT} = D_i + \frac{\hat{e}_i(1-D_i)}{1-\hat{e}_i}$$

Notice that the Soldiers always receive a weight of 1.

IPTW Weights for ATT

Soldiers are always given full weight. "Rare" Civilians are given more weight.



Matching - Summary

- Intuition: For each treated unit, find an untreated unit that "looks" similar. Subtract their outcomes. Average these differences for treatment effect.
- Exact matching: find a control that looks exactly like the treated
- Approximate matching: find a control that looks similar to the treated
- Who "looks alike" can depend on one or more covariates or probability of treatment (i.e. propensity scores)
- We can also weight the treat/control observations as an alternative to a simple DiM estimator