Regression Adjustment

Prof. Mike Aguilar

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

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Regression Adjustment - Preview

- Regression Adjustment (RA) is the application of regression (OLS) to control for a confounder's impact on treatment effect
- OLS for prediction uses "X" to predict "Y"
- OLS for causality uses "X" to estimate $Y_1 - Y_0$
- Use care in choosing "Good" controls
- Combining Matching and Regression can be powerful

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 framework, matching, basic regression analysis

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Does Insurance Improve Health?



Image created with Bing Image Creator
The following is adapted from the book Mastering 'Metrics

Comparing Khuzdar and Maria

Q: Why shouldn't we use Maria to determine impact of insurance on Khuzdar's health?

Khuzdar



VS

Maria



Comparing Khuzdar and Maria

Q: Why shouldn't we use Maria to determine impact of insurance on Khuzdar's health?



A: Maria has more vigor (i.e stronger) than Khuzdar. Comparing their health isn't fair. We don't know if we are capturing the impact of insurance or inherent differences, i.e. there is a selection bias.

Potential Outcomes Notation in Our Case

Y_{1i} : Health for individual i with insurance

Y_{0i} : Health for individual i without insurance

$D_i = 1$: if individual i actually has insurance; 0 if no insurance

$Y_{0i}|D_i = 1$: Khuzdar's health if he didn't have insurance; Y_{0K}

$Y_{0i}|D_i = 0$: Maria's health if she didn't have insurance; Y_{0M}

$Y_{1i}|D_i = 1$: Khuzdar's health if he did have insurance; Y_{1K}

$Y_{1i}|D_i = 0$: Maria's health if she did have insurance; Y_{1M}

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Matching vs Regression

- Matching is a method for overcoming selection bias
- We've already covered Matching, so what's new?

Matching vs Regression

- Matching is a method for overcoming selection bias
- We've already covered Matching, so what's new?
- The goal of this section is to show you that we can replicate Matching via Regression (OLS).

Matching vs Regression

- Matching is a method for overcoming selection bias
- We've already covered Matching, so what's new?
- The goal of this section is to show you that we can replicate Matching via Regression (OLS).
- Seems redundant. Why would you do that?

Matching vs Regression

- Matching is a method for overcoming selection bias
- We've already covered Matching, so what's new?
- The goal of this section is to show you that we can replicate Matching via Regression (OLS).
- Seems redundant. Why would you do that?
- Regression can be applied to cases where Matching is not advisable

Matching vs Regression

- Matching is a method for overcoming selection bias
- We've already covered Matching, so what's new?
- The goal of this section is to show you that we can replicate Matching via Regression (OLS).
- Seems redundant. Why would you do that?
- Regression can be applied to cases where Matching is not advisable
- We will also show that combining Regression WITH Matching can be very powerful

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

- Treated: $E[Y|D = 1]$

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

- Treated: $E[Y|D = 1] = \alpha + \beta \times 1 = \alpha + \beta$
- Untreated: $E[Y|D = 0]$

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

- Treated: $E[Y|D = 1] = \alpha + \beta \times 1 = \alpha + \beta$
- Untreated: $E[Y|D = 0] = \alpha + \beta \times 0 = \alpha$
- Treatment Effect

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

- Treated: $E[Y|D = 1] = \alpha + \beta \times 1 = \alpha + \beta$
- Untreated: $E[Y|D = 0] = \alpha + \beta \times 0 = \alpha$
- Treatment Effect

$$E[Y|D = 1] - E[Y|D = 0] = \alpha + \beta - \alpha = \beta$$

A Linear CEF (Conditional Expectation Function)

- Assume that the conditional expectation of the outcome is a linear function of treatment status.

$$E[Y|D] = \alpha + \beta D$$

- Treated: $E[Y|D = 1] = \alpha + \beta \times 1 = \alpha + \beta$
- Untreated: $E[Y|D = 0] = \alpha + \beta \times 0 = \alpha$
- Treatment Effect

$$E[Y|D = 1] - E[Y|D = 0] = \alpha + \beta - \alpha = \beta$$

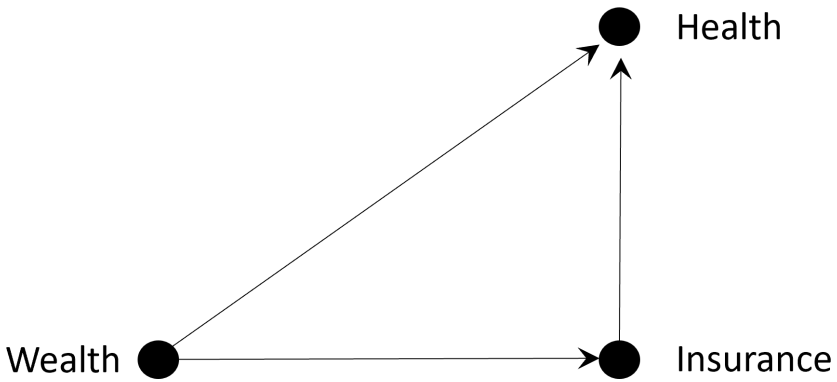
- We can estimate via Difference in Means (DiM)

$$(\bar{Y}|D = 1) - (\bar{Y}|D = 0) = \alpha + \beta - \alpha = \beta$$

Confounders

- The health outcome (Y) may not only be impacted by insurance status (D), but also by other characteristics (J), such as Wealth (measured as continuous $ amount)

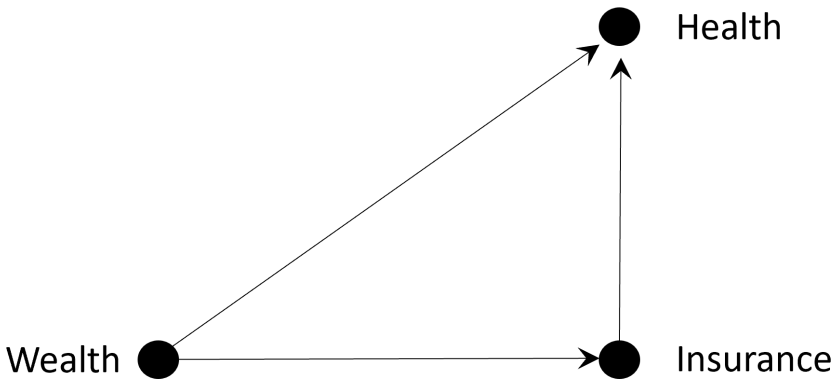
$$E[Y|D, J] = \alpha + \beta D + \gamma J$$



Confounders

- The health outcome (Y) may not only be impacted by insurance status (D), but also by other characteristics (J), such as Wealth (measured as continuous $ amount)

$$E[Y|D, J] = \alpha + \beta D + \gamma J$$

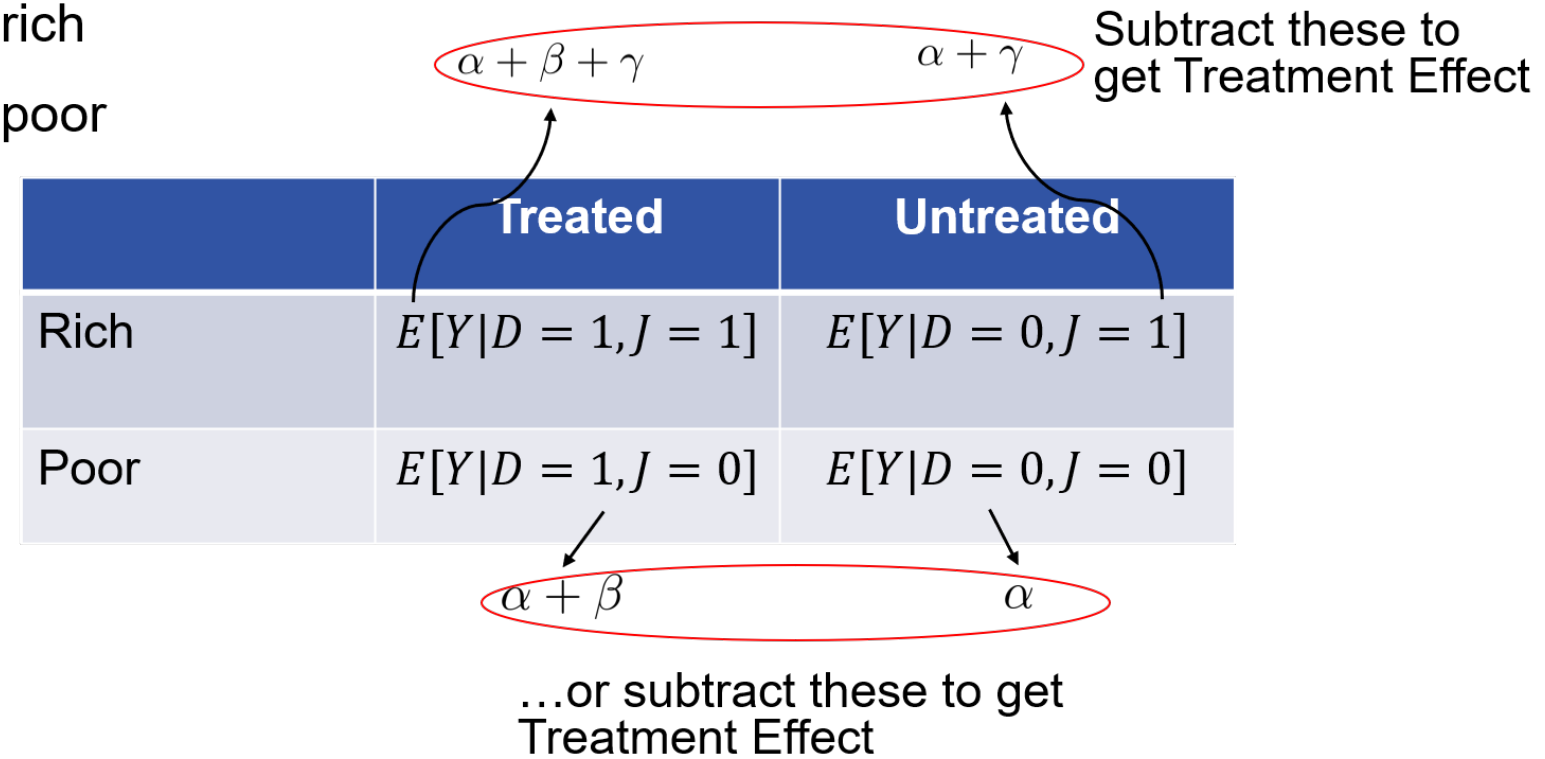


- In Matching we can "control" for the "confounder" by stratifying our sample via levels of J (e.g. converting $ wealth into Rich/Poor binary)

Estimating the Treatment Effect via Covariate Matching

$E[Y|D,J] = \alpha + \beta D + \gamma J$

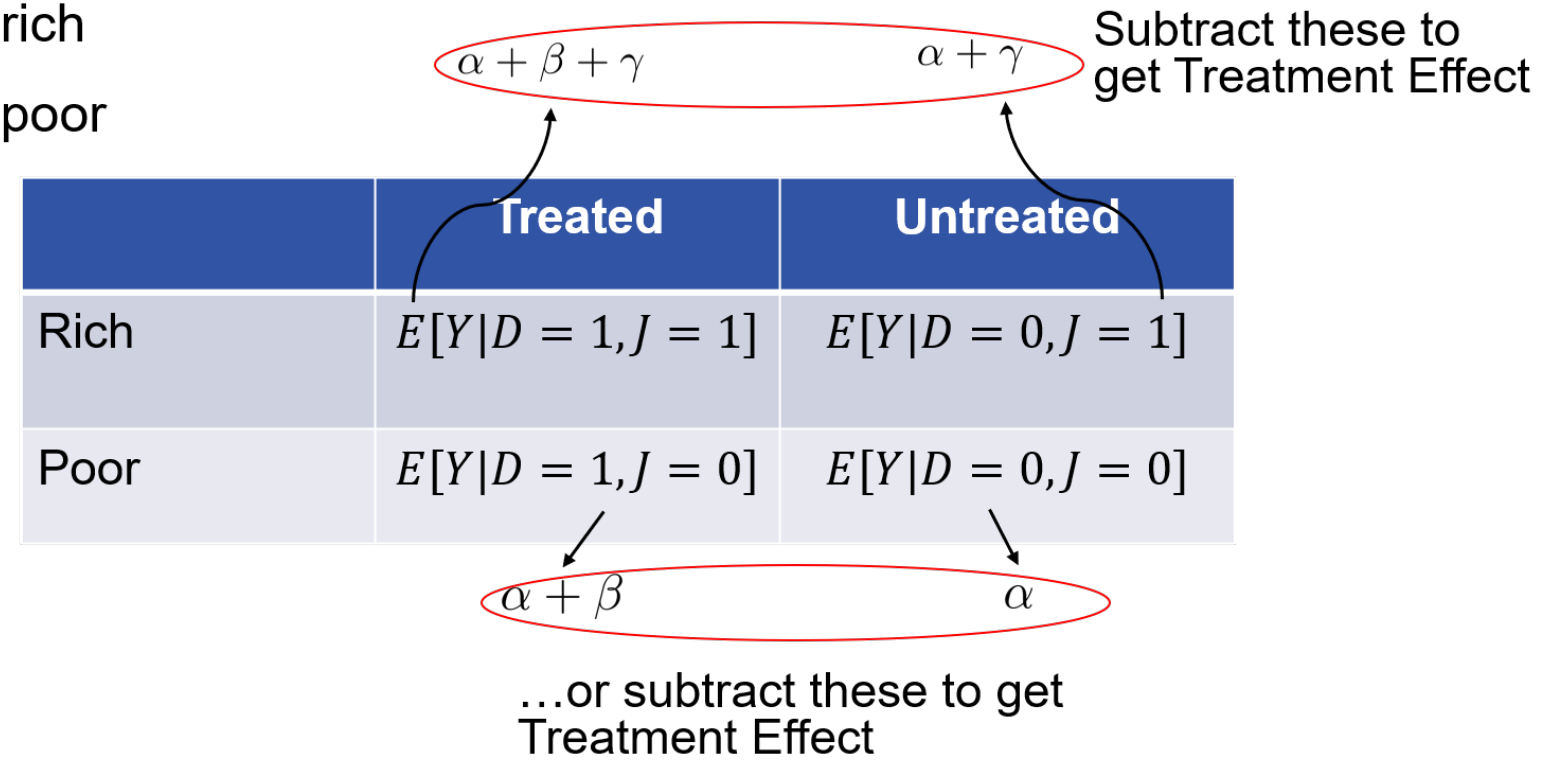
- $J = 1$, if rich
- $J = 0$, if poor



Estimating the Treatment Effect via Covariate Matching

$E[Y|D,J] = \alpha + \beta D + \gamma J$

- $J = 1$, if rich
- $J = 0$, if poor



By conditioning on the controls we are comparing individuals with similar relevant characteristics (i.e. matches)

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...
- where we assume (among other things) $e \perp D$, and so ...

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...
- where we assume (among other things) $e \perp D$, and so ...
- estimating this model by OLS produces "BLUE" estimators of the form...

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...
- where we assume (among other things) $e \perp D$, and so ...
- estimating this model by OLS produces "BLUE" estimators of the form...
- $\hat{\beta}^{OLS} = E[Y|D = 1, J] - E[Y|D = 0, J]$, which is...

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...
- where we assume (among other things) $e \perp D$, and so ...
- estimating this model by OLS produces "BLUE" estimators of the form...
- $\hat{\beta}^{OLS} = E[Y|D = 1, J] - E[Y|D = 0, J]$, which is...$\hat{\beta}^{Matching}$

Regression Recovers the Matching Estimator

- Given the CEF $Y = \alpha + \beta D + \gamma J$...
- we can form the regression model $Y = \alpha + \beta D + \gamma J + e$ ...
- where we assume (among other things) $e \perp D$, and so ...
- estimating this model by OLS produces "BLUE" estimators of the form...
- $\hat{\beta}^{OLS} = E[Y|D = 1, J] - E[Y|D = 0, J]$, which is...$\hat{\beta}^{Matching}$
- OLS recovers the same ATE as Matching*

*Simple OLS may not recover all variants of Matching

Why need regression if Matching works?

Why need regression if Matching works?

- Matching works best for no more than a few discrete variables

Why need regression if Matching works?

- Matching works best for no more than a few discrete variables
- If you have either
  1. many control variables, or

Why need regression if Matching works?

- Matching works best for no more than a few discrete variables
- If you have either
  1. many control variables, or
  2. (near) continuous control variables

Why need regression if Matching works?

- Matching works best for no more than a few discrete variables
- If you have either
  1. many control variables, or
  2. (near) continuous control variables
- Matching becomes impractical (see Matching lecture for details)

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Hierarchy of Causal Settings

1. RCTs
2. Natural Experiments
3. Observational Studies
    - Regression Adjustment (RA)

What's the difference between RA and Regression?

- Regression Adjustment (RA) is Regression (OLS) for the specific purpose of estimating a treatment effect
- Q: A treatment effect is a parameter in the regression model. Isn't all regression about estimating those parameters? Why is this different?

What's the difference between RA and Regression?

- Regression Adjustment (RA) is Regression (OLS) for the specific purpose of estimating a treatment effect
- Q: A treatment effect is a parameter in the regression model. Isn't all regression about estimating those parameters? Why is this different?
- A:
  - We can use regression for prediction of the outcome variable Y. In that setting, we should include all relevant covariates in the regression model.

What's the difference between RA and Regression?

- Regression Adjustment (RA) is Regression (OLS) for the specific purpose of estimating a treatment effect
- Q: A treatment effect is a parameter in the regression model. Isn't all regression about estimating those parameters? Why is this different?
- A:
  - We can use regression for prediction of the outcome variable Y. In that setting, we should include all relevant covariates in the regression model.
  - However, if our focus is on the treatment effect, we should only control for those variables that sharpen our estimate of the associated parameter, i.e. "Good Controls".

What's the difference between RA and Regression?

- Regression Adjustment (RA) is Regression (OLS) for the specific purpose of estimating a treatment effect
- Q: A treatment effect is a parameter in the regression model. Isn't all regression about estimating those parameters? Why is this different?
- A:
  - We can use regression for prediction of the outcome variable Y. In that setting, we should include all relevant covariates in the regression model.
  - However, if our focus is on the treatment effect, we should only control for those variables that sharpen our estimate of the associated parameter, i.e. "Good Controls".
  - Prediction uses "X" to estimate Y
  - Causality uses "X" to estimate $Y_1 - Y_0$

Good vs Bad Controls

Not all controls should be included within your statistical model

Good vs Bad Controls

Not all controls should be included within your statistical model

- Good Control: A Control variable whose inclusion in the statistical model improves the estimate of the treatment effect.

Good vs Bad Controls

Not all controls should be included within your statistical model

- Good Control: A Control variable whose inclusion in the statistical model improves the estimate of the treatment effect.
- Bad Control: A Control variable whose inclusion in the statistical model worsens the estimate of the treatment effect.

Good vs Bad Controls

Not all controls should be included within your statistical model

- Good Control: A Control variable whose inclusion in the statistical model improves the estimate of the treatment effect.
- Bad Control: A Control variable whose inclusion in the statistical model worsens the estimate of the treatment effect.
- Bad Controls violate the orthogonality condition within a linear regression model (i.e. the regressors must be unrelated to the errors). By including Good Controls within the regression model the orthogonality condition is restored.

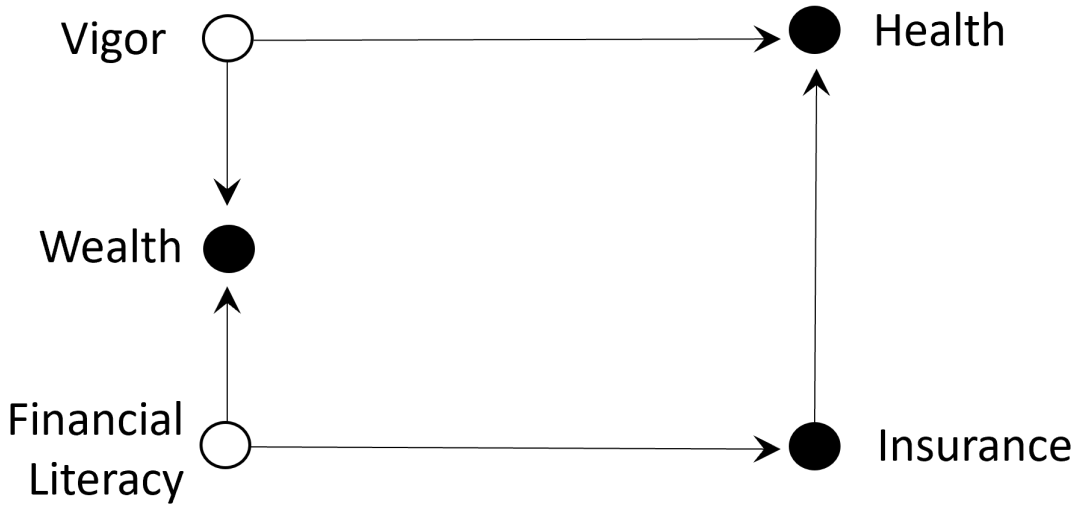
Good vs Bad Controls

Not all controls should be included within your statistical model

- Good Control: A Control variable whose inclusion in the statistical model improves the estimate of the treatment effect.
- Bad Control: A Control variable whose inclusion in the statistical model worsens the estimate of the treatment effect.
- Bad Controls violate the orthogonality condition within a linear regression model (i.e. the regressors must be unrelated to the errors). By including Good Controls within the regression model the orthogonality condition is restored.
- The presence of Bad Controls implies that we shouldn't blindly throw all our data into a model. (WARNING: Avoid the ML-temptation to use all of your variables and rely on dimension reduction)

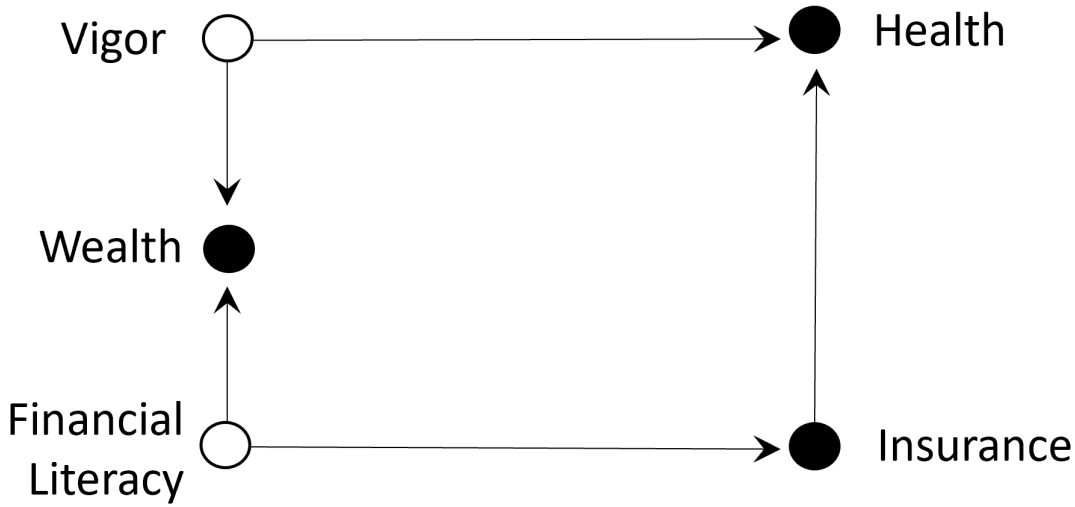
Multiple Confounders

Q: Can you explain this DAG?



Multiple Confounders

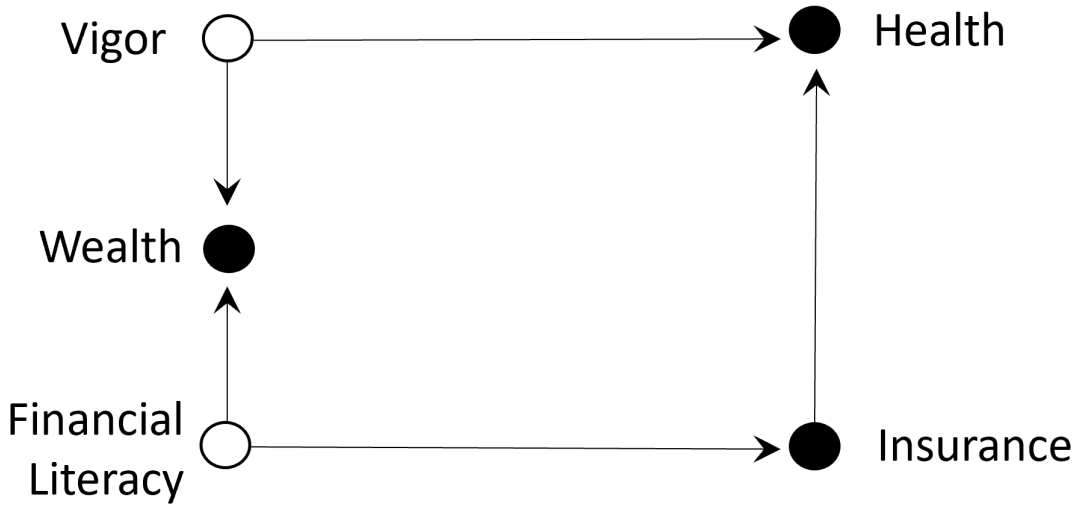
Q: Can you explain this DAG?



A: The black dots represent observed data, while the white dots represent unobserved data. The story here is that vigor (e.g. Khuzdar gets sick often while Maria does not) may impact health and wealth

Multiple Confounders

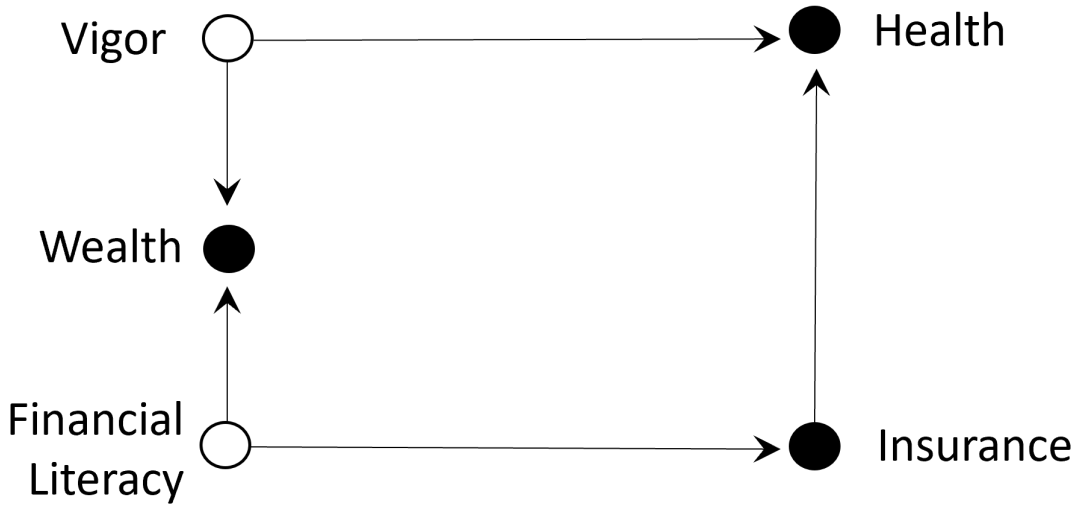
Q: Can you explain this DAG?



A: The black dots represent observed data, while the white dots represent unobserved data. The story here is that vigor (e.g. Khuzdar gets sick often while Maria does not) may impact health and wealth (people with high vigor have more energy to work, etc..).

Multiple Confounders

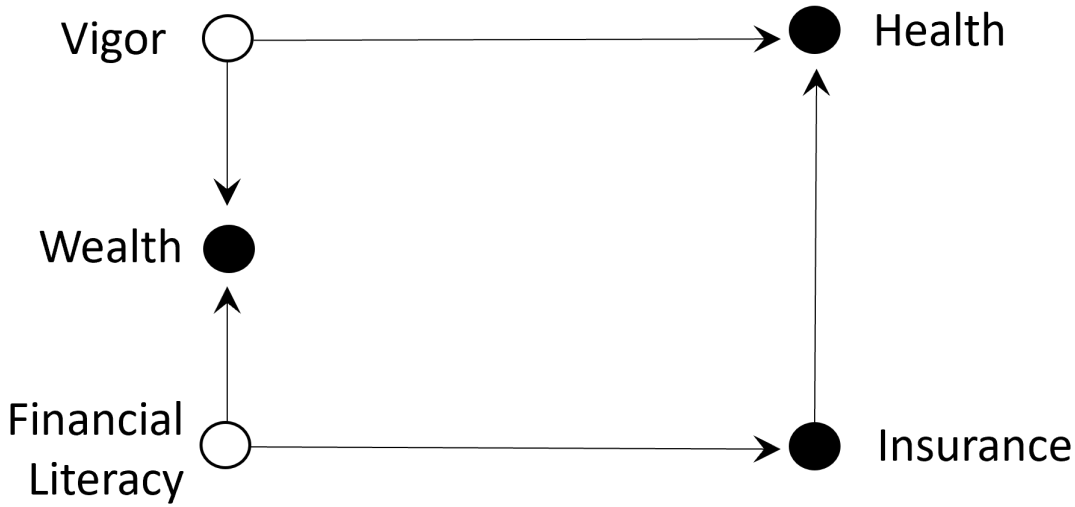
Q: Can you explain this DAG?



A: The black dots represent observed data, while the white dots represent unobserved data. The story here is that vigor (e.g. Khuzdar gets sick often while Maria does not) may impact health and wealth (people with high vigor have more energy to work, etc..). Moreover, financially literate people might be able to accumulate more wealth,

Multiple Confounders

Q: Can you explain this DAG?



A: The black dots represent observed data, while the white dots represent unobserved data. The story here is that vigor (e.g. Khuzdar gets sick often while Maria does not) may impact health and wealth (people with high vigor have more energy to work, etc..). Moreover, financially literate people might be able to accumulate more wealth, and might be more likely to accept jobs with health insurance.

Your Regression Is Not What You Think

Suppose the Data Generating Process (DGP) is

$$Health = \alpha + \beta Ins(Fin.Lit) + Vigor$$

where $Ins(Fin.Lit)$ can be read to mean that having Insurance is a function of Financial Literacy.

Your Regression Is Not What You Think

Suppose the Data Generating Process (DGP) is

$$Health = \alpha + \beta Ins(Fin.Lit) + Vigor$$

where $Ins(Fin.Lit)$ can be read to mean that having Insurance is a function of Financial Literacy. Suppose we control for Wealth by running:

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

Your Regression Is Not What You Think

Suppose the Data Generating Process (DGP) is

$$Health = \alpha + \beta Ins(Fin.Lit) + Vigor$$

where $Ins(Fin.Lit)$ can be read to mean that having Insurance is a function of Financial Literacy. Suppose we control for Wealth by running:

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

We're actually running:

$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor)$$

Don't Include Bad Controls

$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor) \quad (1)$$

Q: What's the issue caused by controlling for wealth?

Don't Include Bad Controls

$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor) \quad (1)$$

Q: What's the issue caused by controlling for wealth?
A:

- Error is related to a regressor

Don't Include Bad Controls

$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor) \quad (1)$$

Q: What's the issue caused by controlling for wealth?
A:

- Error is related to a regressor
- $e \not\perp Wealth$ because $Wealth \leftarrow Vigor \rightarrow e$

Don't Include Bad Controls

$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor) \quad (1)$$

Q: What's the issue caused by controlling for wealth?
A:

- Error is related to a regressor
- $e \not\perp Wealth$ because $Wealth \leftarrow Vigor \rightarrow e$
- Wealth is a collider (BAD CONTROL)

Don't Include Bad Controls

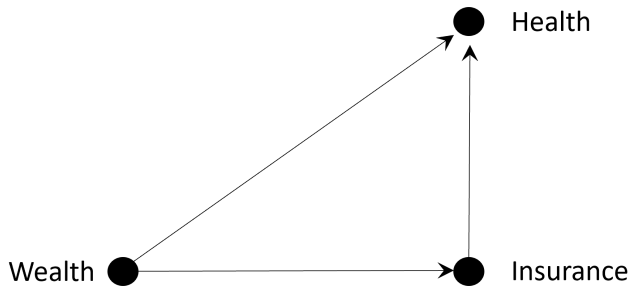
$$Health = \alpha + \beta Ins(Fin.Lit) + \gamma Wealth(Fin.Lit, Vigor) + e(Vigor) \quad (1)$$

Q: What's the issue caused by controlling for wealth?
A:

- Error is related to a regressor
- $e \not\perp Wealth$ because $Wealth \leftarrow Vigor \rightarrow e$
- Wealth is a collider (BAD CONTROL)
- It is better NOT to include Wealth in your estimate

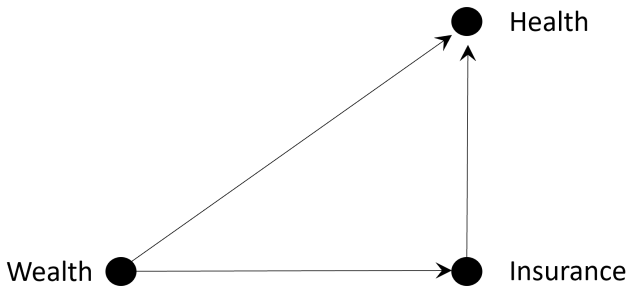
Your Regression Is Not What you Think

Suppose the DGP is $Health = \alpha + \beta Ins + \gamma Wealth$



Your Regression Is Not What you Think

Suppose the DGP is $Health = \alpha + \beta Ins + \gamma Wealth$

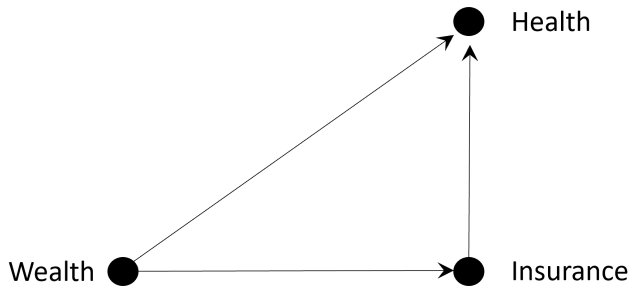


What if we mistakenly don't control for Wealth. We think we're estimating:

$$Health = \alpha + \beta Ins + e$$

Your Regression Is Not What you Think

Suppose the DGP is $Health = \alpha + \beta Ins + \gamma Wealth$



What if we mistakenly don't control for Wealth. We think we're estimating:

$$Health = \alpha + \beta Ins + e$$

But we're actually estimating

$$Health = \alpha + \beta Ins.(Wealth) + e(Wealth)$$

where $e(Wealth)$ implies that error includes the omitted Wealth.

Note: We can include errors in DAGs as unobservable. We'll do that when we cover Omitted Variable Bias.

Include Good Controls

Q: Why is running this regression bad?

$$Health = \alpha + \beta Ins.(Wealth) + e(Wealth)$$

Include Good Controls

Q: Why is running this regression bad?

$$Health = \alpha + \beta Ins.(Wealth) + e(Wealth)$$

A:

- Error is related to a regressor
- $e \not\perp Ins$ because $Ins \leftarrow Wealth \rightarrow e$
- Wealth is a confounder (GOOD CONTROL)
- We should estimate $Health = \alpha + \beta Ins.(Wealth) + \gamma Wealth + e$

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?
A:

- We don't KNOW

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?
A:

- We don't KNOW
- We have to ASSUME

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?
A:

- We don't KNOW
- We have to ASSUME
- The DAG is drawn as a proxy for the DGP BEFORE estimation

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?
A:

- We don't KNOW
- We have to ASSUME
- The DAG is drawn as a proxy for the DGP BEFORE estimation
- Subject matter expertise helps us form the DAG

Identifying Good and Bad Controls

Q: How do we know if a confounder is a Good or Bad control?
A:

- We don't KNOW
- We have to ASSUME
- The DAG is drawn as a proxy for the DGP BEFORE estimation
- Subject matter expertise helps us form the DAG
- We then test if the data is consistent with our assumptions

More on Good and Bad Controls

- See lecture notes on DAGs for more details
- The notion of Good and Bad Controls permeates all observational studies
- We didn't discuss these much in Matching since the scope of variables that can be included in Matching is rather limited

More on Good and Bad Controls

- See lecture notes on DAGs for more details
- The notion of Good and Bad Controls permeates all observational studies
- We didn't discuss these much in Matching since the scope of variables that can be included in Matching is rather limited
- In RA, there is a temptation to control for everything in your dataset. After all, more control variables increases the likelihood that we can overcome selection bias.

More on Good and Bad Controls

- See lecture notes on DAGs for more details
- The notion of Good and Bad Controls permeates all observational studies
- We didn't discuss these much in Matching since the scope of variables that can be included in Matching is rather limited
- In RA, there is a temptation to control for everything in your dataset. After all, more control variables increases the likelihood that we can overcome selection bias.
- But, ...
  1. lots of covariates in a regression can lower precision, and...

More on Good and Bad Controls

- See lecture notes on DAGs for more details
- The notion of Good and Bad Controls permeates all observational studies
- We didn't discuss these much in Matching since the scope of variables that can be included in Matching is rather limited
- In RA, there is a temptation to control for everything in your dataset. After all, more control variables increases the likelihood that we can overcome selection bias.
- But, ...
  1. lots of covariates in a regression can lower precision, and...
  2. using lots of covariates increases the chances that you are going to accidentally include a Bad Control.

Automatic Control Selection

Double Selection

Some ML enthusiasts support a Double Selection approach to identify good/bad controls

Automatic Control Selection

Double Selection

Some ML enthusiasts support a Double Selection approach to identify good/bad controls

1. Estimate LASSO($Y \sim X$) and extract subset $\tilde{X}^Y$ that are good predictors of outcome Y

Automatic Control Selection

Double Selection

Some ML enthusiasts support a Double Selection approach to identify good/bad controls

1. Estimate LASSO($Y \sim X$) and extract subset $\tilde{X}^Y$ that are good predictors of outcome Y
2. Estimate LASSO($D \sim X$) and extract subset $\tilde{X}^D$ that are good predictors of treatment D

Automatic Control Selection

Double Selection

Some ML enthusiasts support a Double Selection approach to identify good/bad controls

1. Estimate LASSO($Y \sim X$) and extract subset $\tilde{X}^Y$ that are good predictors of outcome Y
2. Estimate LASSO($D \sim X$) and extract subset $\tilde{X}^D$ that are good predictors of treatment D
3. Estimate OLS($Y \sim D + \tilde{X}^Y + \tilde{X}^D$) and extract the treatment effect via the coefficient on D

Automatic Control Selection

Double Selection

Some ML enthusiasts support a Double Selection approach to identify good/bad controls

1. Estimate LASSO($Y \sim X$) and extract subset $\tilde{X}^Y$ that are good predictors of outcome Y
2. Estimate LASSO($D \sim X$) and extract subset $\tilde{X}^D$ that are good predictors of treatment D
3. Estimate OLS($Y \sim D + \tilde{X}^Y + \tilde{X}^D$) and extract the treatment effect via the coefficient on D

Hundermund, et al (2023) warn against this approach in most applied settings due to risk of colliders or other bad controls remaining among the subsets of X.
However, we could still use Double Selection on a large set of "Good Controls".

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

- ATT: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

- ATT: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$
- Selection Bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

- ATT: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$
- Selection Bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$

Q: Suppose only the "rich" people have health insurance. How might that impact the Selection Bias?

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

- ATT: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$
- Selection Bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$

Q: Suppose only the "rich" people have health insurance. How might that impact the Selection Bias?
A:

- We might expect $E[Y_{0i}|D_i = 1, Rich] > E[Y_{0i}|D_i = 0, Poor]$ such that the rich are healthier even without insurance due to access to better quality of food, etc..

Selection Bias in Health Insurance Case Study

Potential Outcomes Perspective

Recall that $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is comprised of two pieces

- ATT: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$
- Selection Bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$

Q: Suppose only the "rich" people have health insurance. How might that impact the Selection Bias?
A:

- We might expect $E[Y_{0i}|D_i = 1, Rich] > E[Y_{0i}|D_i = 0, Poor]$ such that the rich are healthier even without insurance due to access to better quality of food, etc..
- This systematic difference in our treated and control groups implies that we shouldn't estimate the treatment effect using only the observed levels of health.

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$E[Y|D = 1, Z = z] =$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$E[Y|D = 1, Z = z] = \alpha + \beta + \gamma z + E[e|D = 1, Z = z]$$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$E[Y|D = 1, Z = z] = \alpha + \beta + \gamma z + E[e|D = 1, Z = z]$$
$$E[Y|D = 0, Z = z] =$$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1, Z=z] =& \alpha + \beta + \gamma z + E[e|D=1, Z=z] \\ E[Y|D=0, Z=z] =& \alpha + \gamma z + E[e|D=0, Z=z] \end{aligned}$$

The ATE is

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1, Z=z] =& \alpha + \beta + \gamma z + E[e|D=1, Z=z] \\ E[Y|D=0, Z=z] =& \alpha + \gamma z + E[e|D=0, Z=z] \end{aligned}$$

The ATE is β.

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1,Z=z] =& \alpha + \beta + \gamma z + E[e|D=1,Z=z] \\ E[Y|D=0,Z=z] =& \alpha + \gamma z + E[e|D=0,Z=z] \end{aligned}$$

The ATE is β. This is derived via

$$\begin{aligned} E[Y|D=1,Z=z] - E[Y|D=0,Z=z] = \\ \beta + E[e|D=1,Z=z] - E[e|D=0,Z=z] \end{aligned}$$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1,Z=z] =& \alpha + \beta + \gamma z + E[e|D=1,Z=z] \\ E[Y|D=0,Z=z] =& \alpha + \gamma z + E[e|D=0,Z=z] \end{aligned}$$

The ATE is β. This is derived via

$$\begin{aligned} E[Y|D=1,Z=z] - E[Y|D=0,Z=z] = \\ \beta + E[e|D=1,Z=z] - E[e|D=0,Z=z] \end{aligned}$$

- $e \perp \{D, Z\} \Rightarrow$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1,Z=z] =& \alpha + \beta + \gamma z + E[e|D=1,Z=z] \\ E[Y|D=0,Z=z] =& \alpha + \gamma z + E[e|D=0,Z=z] \end{aligned}$$

The ATE is β. This is derived via

$$\begin{aligned} E[Y|D=1,Z=z] - E[Y|D=0,Z=z] = \\ \beta + E[e|D=1,Z=z] - E[e|D=0,Z=z] \end{aligned}$$

- $e \perp \{D, Z\} \Rightarrow$
- $E[e|D=1,Z=z] = E[e|D=0,Z=z] \Rightarrow$

Selection bias in RA

Orthogonality of Errors and Regressors overcomes Selection Bias in RA

Consider the following population regression model

$$Y = \alpha + \beta D + \gamma Z + e$$

$$\begin{aligned} E[Y|D=1,Z=z] =& \alpha + \beta + \gamma z + E[e|D=1,Z=z] \\ E[Y|D=0,Z=z] =& \alpha + \gamma z + E[e|D=0,Z=z] \end{aligned}$$

The ATE is β. This is derived via

$$\begin{aligned} E[Y|D=1,Z=z] - E[Y|D=0,Z=z] = \\ \beta + E[e|D=1,Z=z] - E[e|D=0,Z=z] \end{aligned}$$

- $e \perp \{D, Z\} \Rightarrow$
- $E[e|D=1,Z=z] = E[e|D=0,Z=z] \Rightarrow$
- No Selection Bias

Selection Bias in our Case Study

Regression Perspective

Suppose we (mistakenly) estimated

$$Health = \alpha + \beta Ins + e$$

Selection Bias in our Case Study

Regression Perspective

Suppose we (mistakenly) estimated

$$Health = \alpha + \beta Ins + e$$

We stated previously that Wealth is a confounder, but since it's omitted our model looks like

$$Health = \alpha + \beta Ins + e(Wealth)$$

i.e. Wealth resides within the error.

Selection Bias in our Case Study

Regression Perspective

Suppose we (mistakenly) estimated

$$Health = \alpha + \beta Ins + e$$

We stated previously that Wealth is a confounder, but since it's omitted our model looks like

$$Health = \alpha + \beta Ins + e(Wealth)$$

i.e. Wealth resides within the error. So when we estimate via OLS

$$\hat{\beta} = \beta + [e(Wealth)|Ins = 1] - [e(Wealth)|Ins = 0)]$$

Selection Bias in our Case Study

Regression Perspective

Suppose we (mistakenly) estimated

$$Health = \alpha + \beta Ins + e$$

We stated previously that Wealth is a confounder, but since it's omitted our model looks like

$$Health = \alpha + \beta Ins + e(Wealth)$$

i.e. Wealth resides within the error. So when we estimate via OLS

$$\hat{\beta} = \beta + [e(Wealth)|Ins = 1] - [e(Wealth)|Ins = 0)]$$

Unless the last term is 0, our estimator is biased.

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise.

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.
- We're concerned that the insured have different omitted determinants of health than the uninsured, $E[e|Ins = 1] \neq E[e|Ins = 0]$, because ...

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.
- We're concerned that the insured have different omitted determinants of health than the uninsured, $E[e|Ins = 1] \neq E[e|Ins = 0]$, because ...
- ... (in an extreme setting) the insured are Rich and the uninsured are Poor, and ...

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.
- We're concerned that the insured have different omitted determinants of health than the uninsured, $E[e|Ins = 1] \neq E[e|Ins = 0]$, because ...
- ... (in an extreme setting) the insured are Rich and the uninsured are Poor, and ...
- ... being Rich/Poor can directly impact health, so ...

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.
- We're concerned that the insured have different omitted determinants of health than the uninsured, $E[e|Ins = 1] \neq E[e|Ins = 0]$, because ...
- ... (in an extreme setting) the insured are Rich and the uninsured are Poor, and ...
- ... being Rich/Poor can directly impact health, so ...
- ... our regression errors are correlated with the regressor (Insurance), implying ...

Selection Bias is an Orthogonality Violation

Q: What does $[e(\text{Wealth})|Ins = 1] - [e(\text{Wealth})|Ins = 0]$ mean?
A:

- The errors are the portion of health that is not explained by insurance, which may include systematic components due to omitted variables as well as noise. Note: we want the errors to be only noise
- This term is the difference in the average omitted components of health between people with insurance and those without.
- We're concerned that the insured have different omitted determinants of health than the uninsured, $E[e|Ins = 1] \neq E[e|Ins = 0]$, because ...
- ... (in an extreme setting) the insured are Rich and the uninsured are Poor, and ...
- ... being Rich/Poor can directly impact health, so ...
- ... our regression errors are correlated with the regressor (Insurance), implying ...
- ... the orthogonality condition is violated, which is equivalent to selection bias.

Resolving Selection Bias with RA

If we instead control for Wealth in our regression (Regression Adjustment)

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

Resolving Selection Bias with RA

If we instead control for Wealth in our regression (Regression Adjustment)

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

...then our estimator

$$\hat{\beta} = \beta + [e|Ins = 1, Wealth] - [e|Ins = 0, Wealth]$$

is unbiased since $e \perp Ins|Wealth$.

Resolving Selection Bias with RA

If we instead control for Wealth in our regression (Regression Adjustment)

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

...then our estimator

$$\hat{\beta} = \beta + [e|Ins = 1, Wealth] - [e|Ins = 0, Wealth]$$

is unbiased since $e \perp Ins|Wealth$.
Q: What does this orthogonality mean in our case?

Resolving Selection Bias with RA

If we instead control for Wealth in our regression (Regression Adjustment)

$$Health = \alpha + \beta Ins + \gamma Wealth + e$$

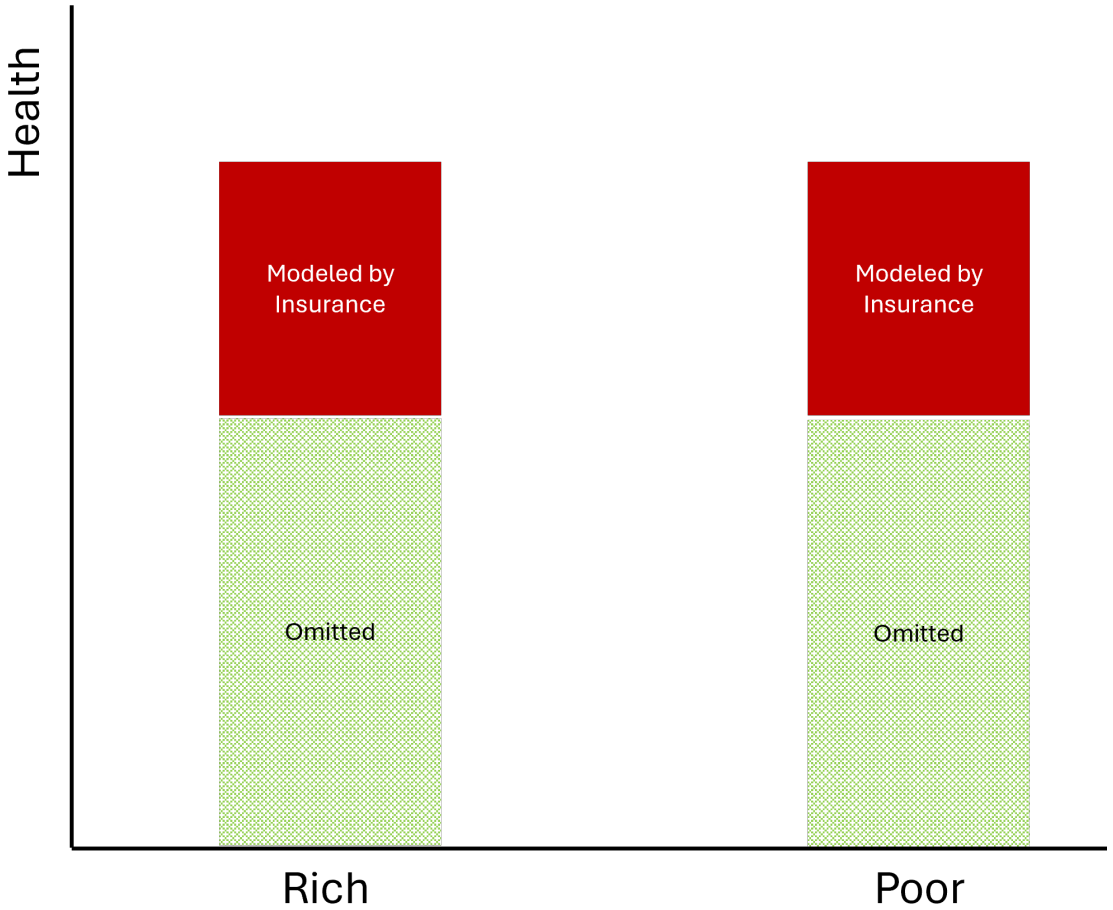
...then our estimator

$$\hat{\beta} = \beta + [e|Ins = 1, Wealth] - [e|Ins = 0, Wealth]$$

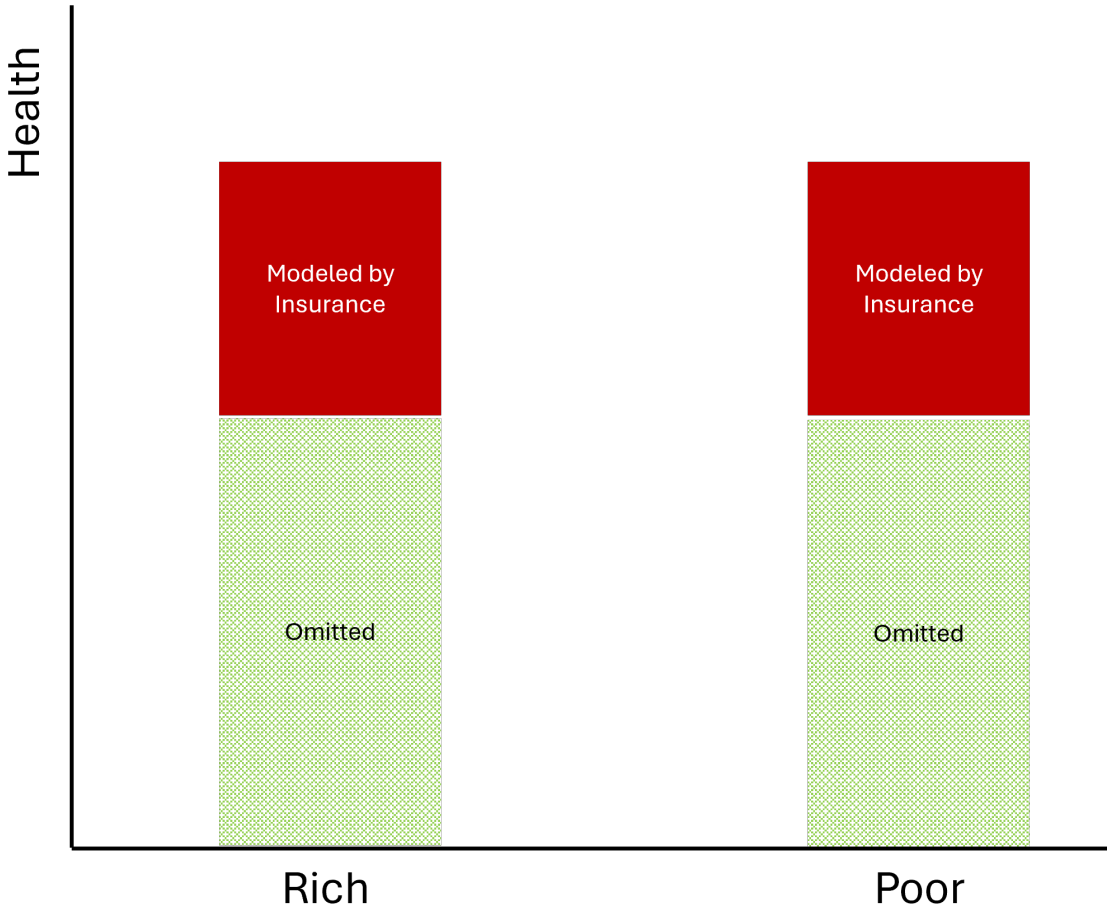
is unbiased since $e \perp Ins|Wealth$.
Q: What does this orthogonality mean in our case?
A: The unexplained portions of health are unrelated across the treated and control groups once we control for wealth.

Visualizing the error and orthogonality

Visualizing the error and orthogonality

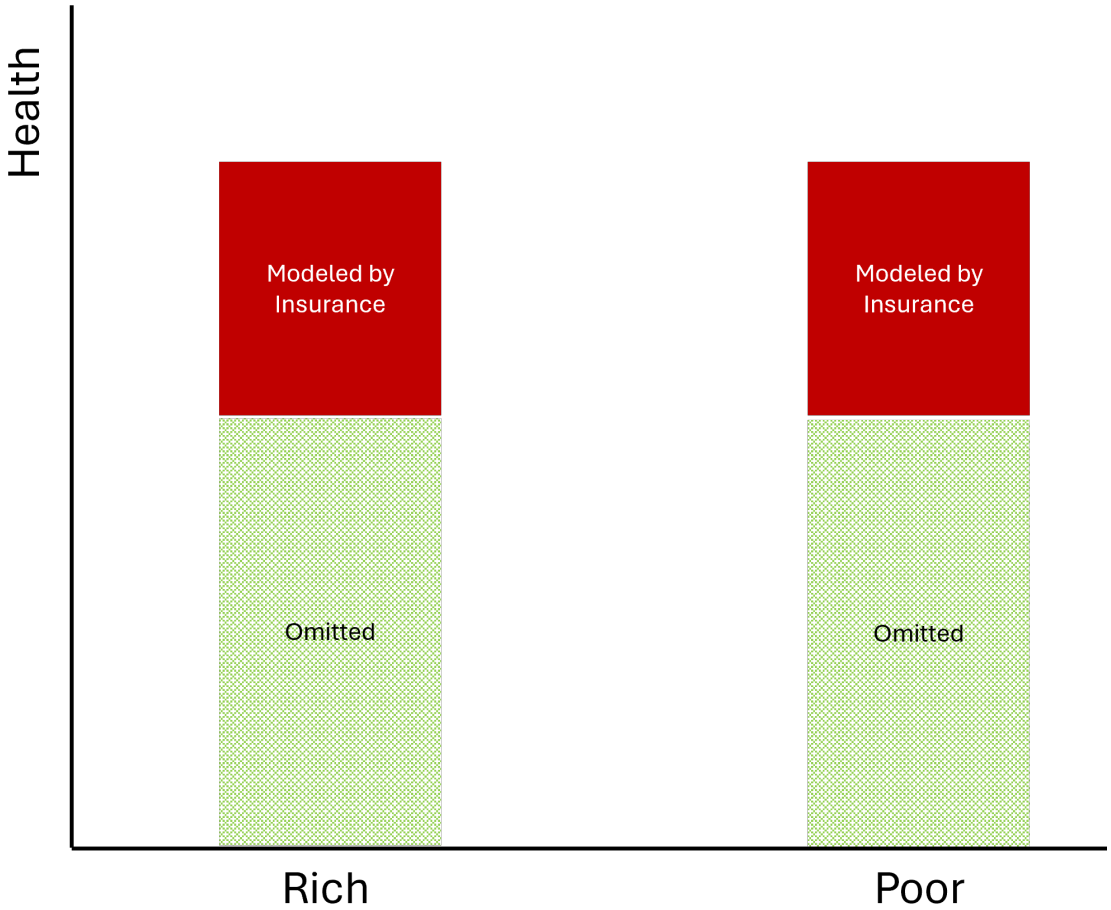


Visualizing the error and orthogonality



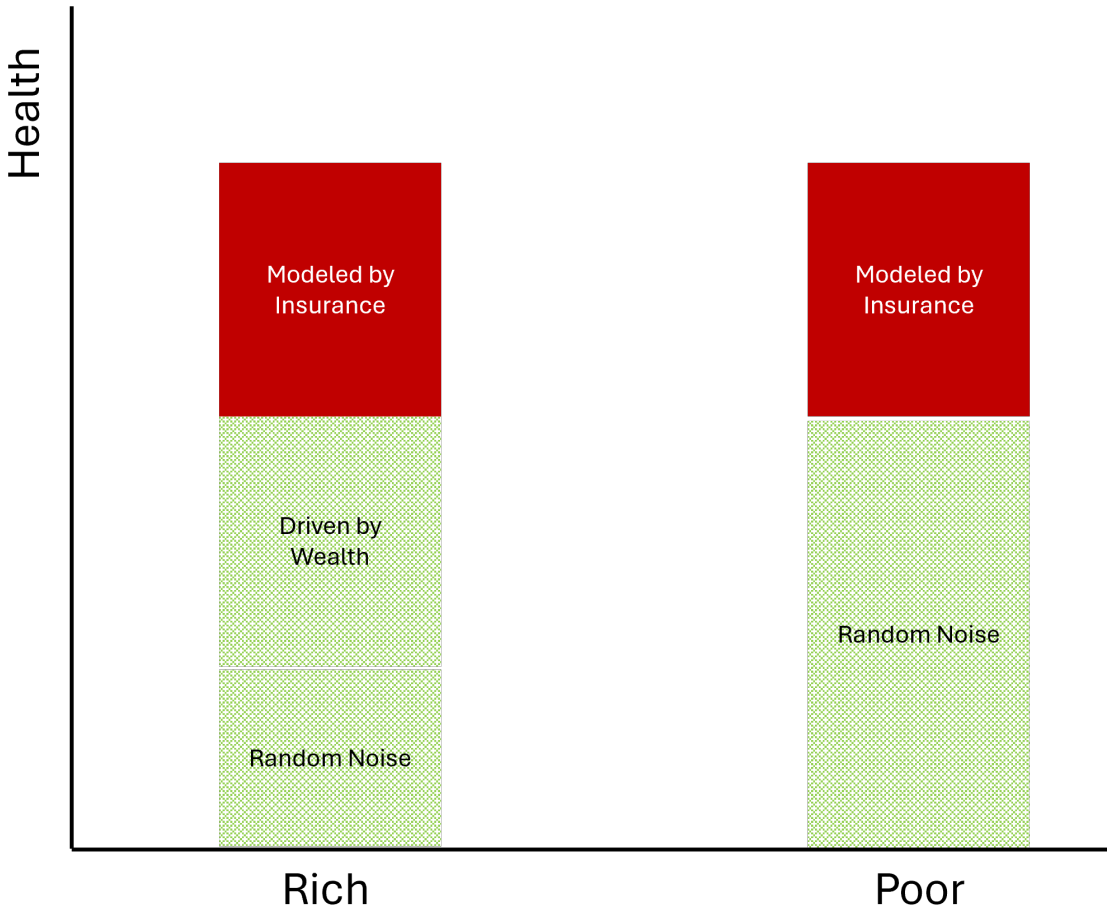
- We're NOT saying that Health varies for Rich or Poor (which it might)

Visualizing the error and orthogonality



- We're NOT saying that Health varies for Rich or Poor (which it might)
- We're NOT saying that insurance explains more for Rich than Poor (which it might)...

Visualizing the error and orthogonality



- We're saying

Visualizing the error and orthogonality



- We're saying that the omitted determinants of health differ structurally for Rich and Poor...

Visualizing the error and orthogonality



- We're saying that the omitted determinants of health differ structurally for Rich and Poor...
- .... implying that the regressor error is correlated with insurance status

Visualizing the error and orthogonality

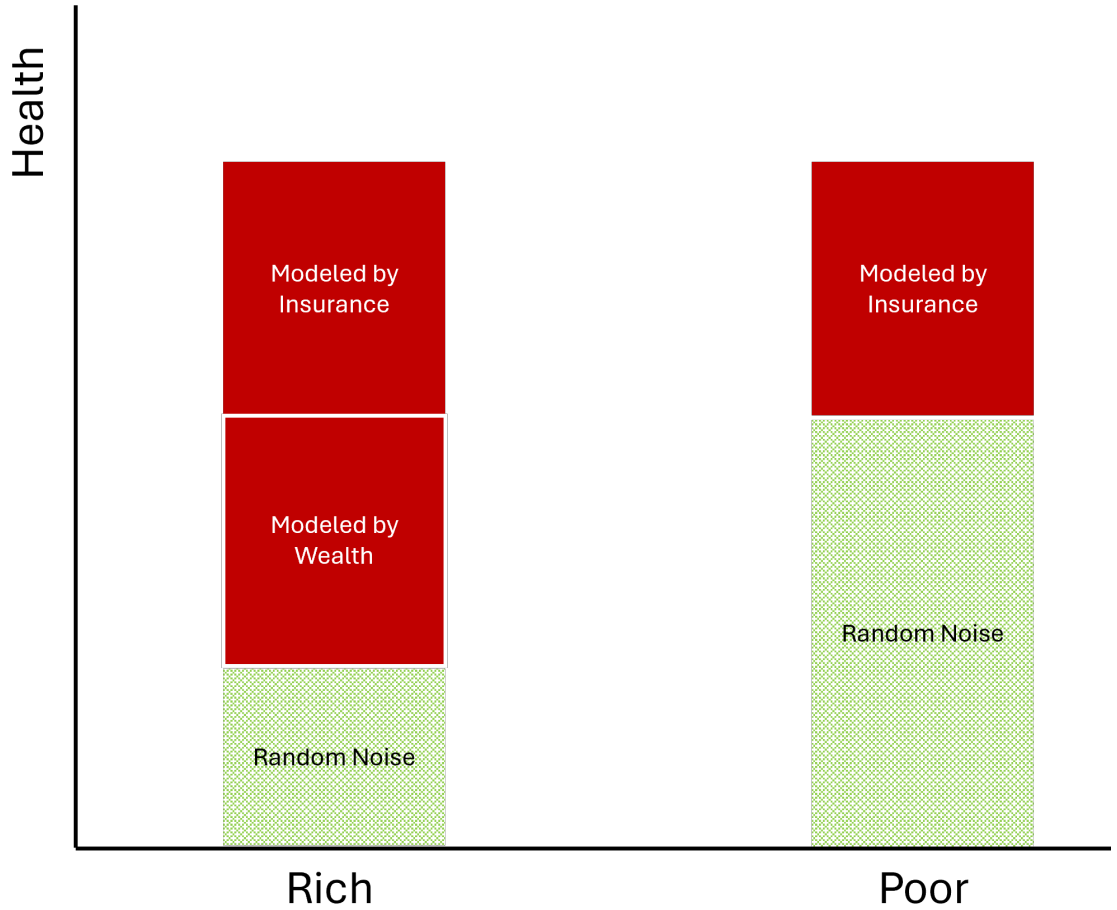


Figure: Caption

- But once we control for Wealth,

Visualizing the error and orthogonality



Figure: Caption

- But once we control for Wealth, the remaining error is no longer systematically related to insurance status

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

Combining Matching and Regression

Implementation is a simple extension of the tools you've already developed

Combining Matching and Regression

Implementation is a simple extension of the tools you've already developed

1. Generate matches for the treated and control groups. You can match on one or several covariates, impose calipers, etc..

Combining Matching and Regression

Implementation is a simple extension of the tools you've already developed

1. Generate matches for the treated and control groups. You can match on one or several covariates, impose calipers, etc..
2. Run a regression on the matched sample only

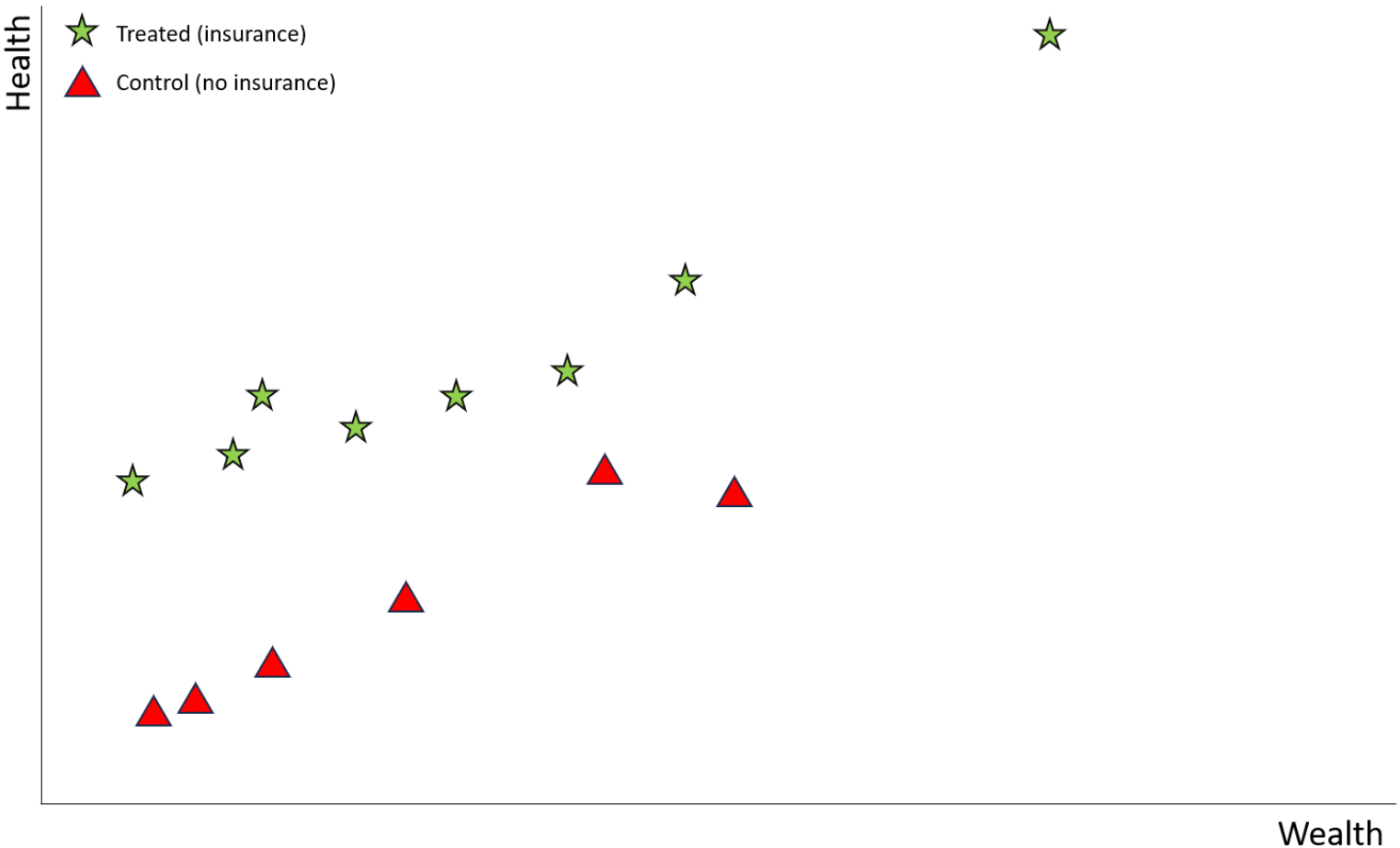
Combining Matching and Regression

Implementation is a simple extension of the tools you've already developed

1. Generate matches for the treated and control groups. You can match on one or several covariates, impose calipers, etc..
2. Run a regression on the matched sample only

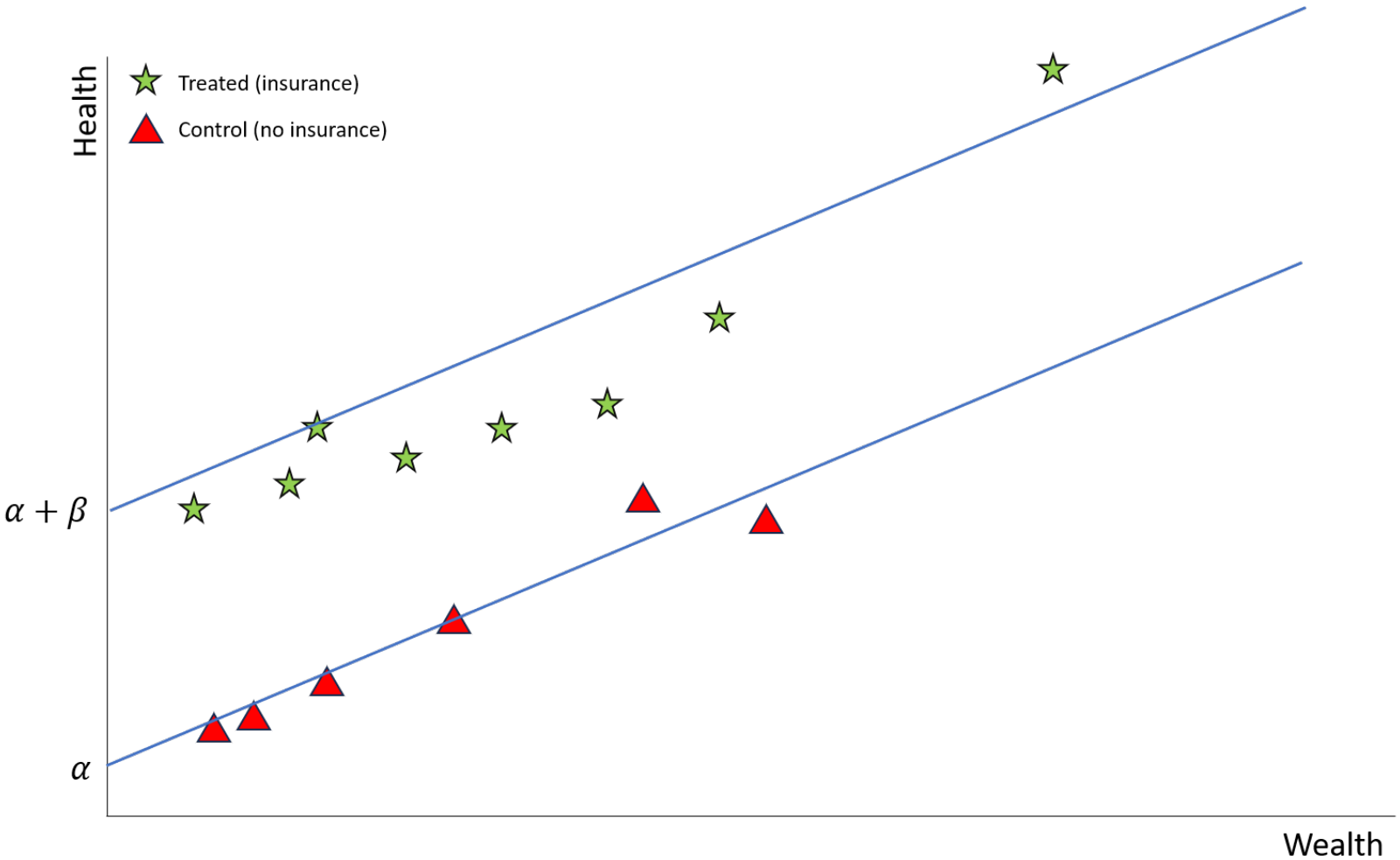
Combining matching and regression in this manner has been shown in practice to improve bias/precision.

Consider some data



Estimate a regression using all the data

$\widehat{ATE} = \hat{\beta}$ from $health = \alpha + \beta Ins + \gamma wealth$

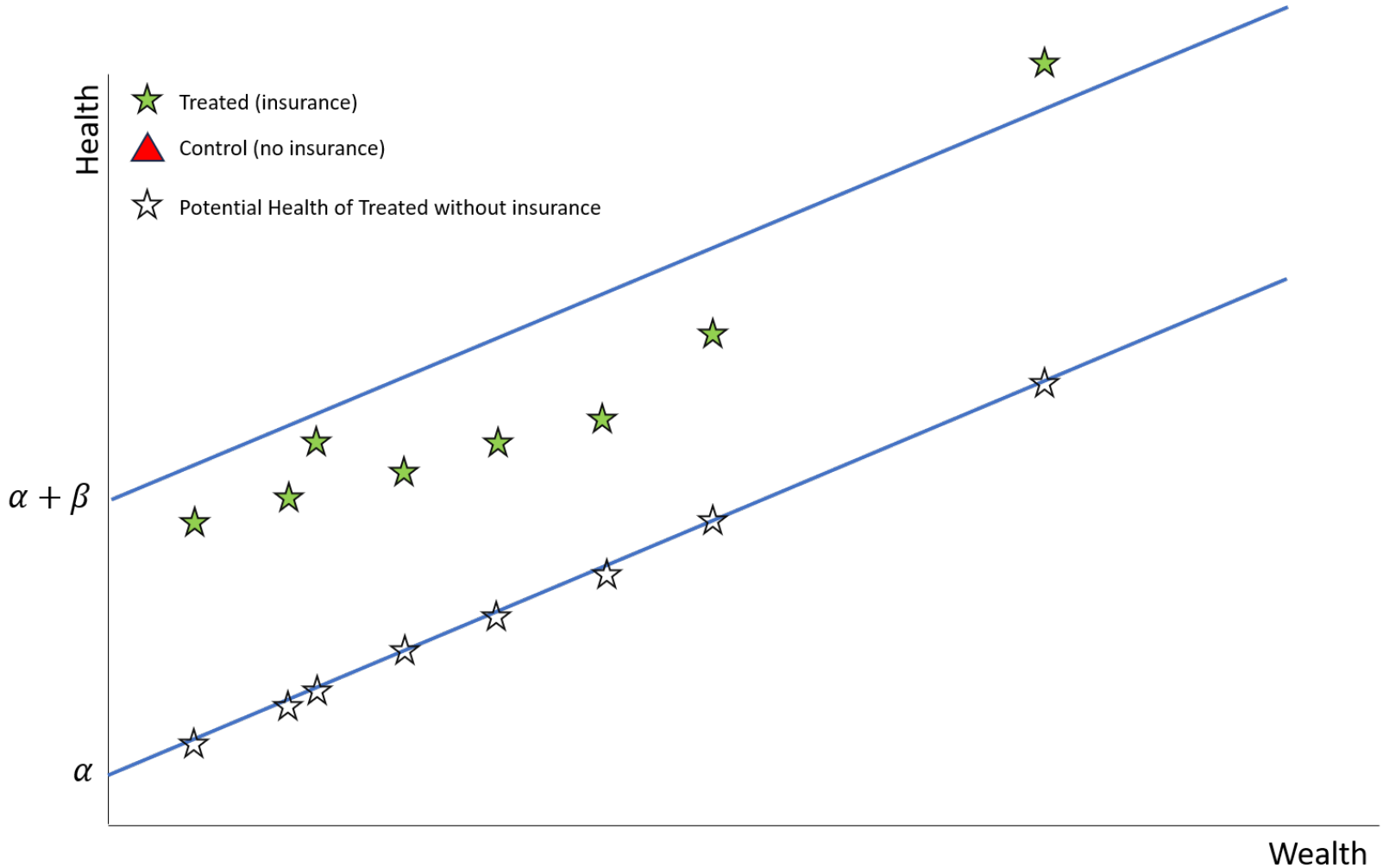


Recovering Potential Outcomes

Recall that $ATE = E[health|ins = 1, wealth] - E[health|ins = 0, wealth]$ relies on difference in potential outcomes. OLS uses the information from the control group to estimate the unobserved outcomes; e.g.what health would have been for the treated group if they didn't have insurance.

Recovering Potential Outcomes

Recall that $ATE = E[health|ins = 1, wealth] - E[health|ins = 0, wealth]$ relies on difference in potential outcomes. OLS uses the information from the control group to estimate the unobserved outcomes; e.g.what health would have been for the treated group if they didn't have insurance.



The Problem of Extrapolating Beyond Common Support

Q: Why does the treated unit with the highest wealth pose a problem?

The Problem of Extrapolating Beyond Common Support

Q: Why does the treated unit with the highest wealth pose a problem?

A:

- There are no control units with such high wealth in our dataset

The Problem of Extrapolating Beyond Common Support

Q: Why does the treated unit with the highest wealth pose a problem?

A:

- There are no control units with such high wealth in our dataset
- We are extrapolating beyond on our common support

The Problem of Extrapolating Beyond Common Support

Q: Why does the treated unit with the highest wealth pose a problem?

A:

- There are no control units with such high wealth in our dataset
- We are extrapolating beyond on our common support

Q: So what's a solution?

The Problem of Extrapolating Beyond Common Support

Q: Why does the treated unit with the highest wealth pose a problem?

A:

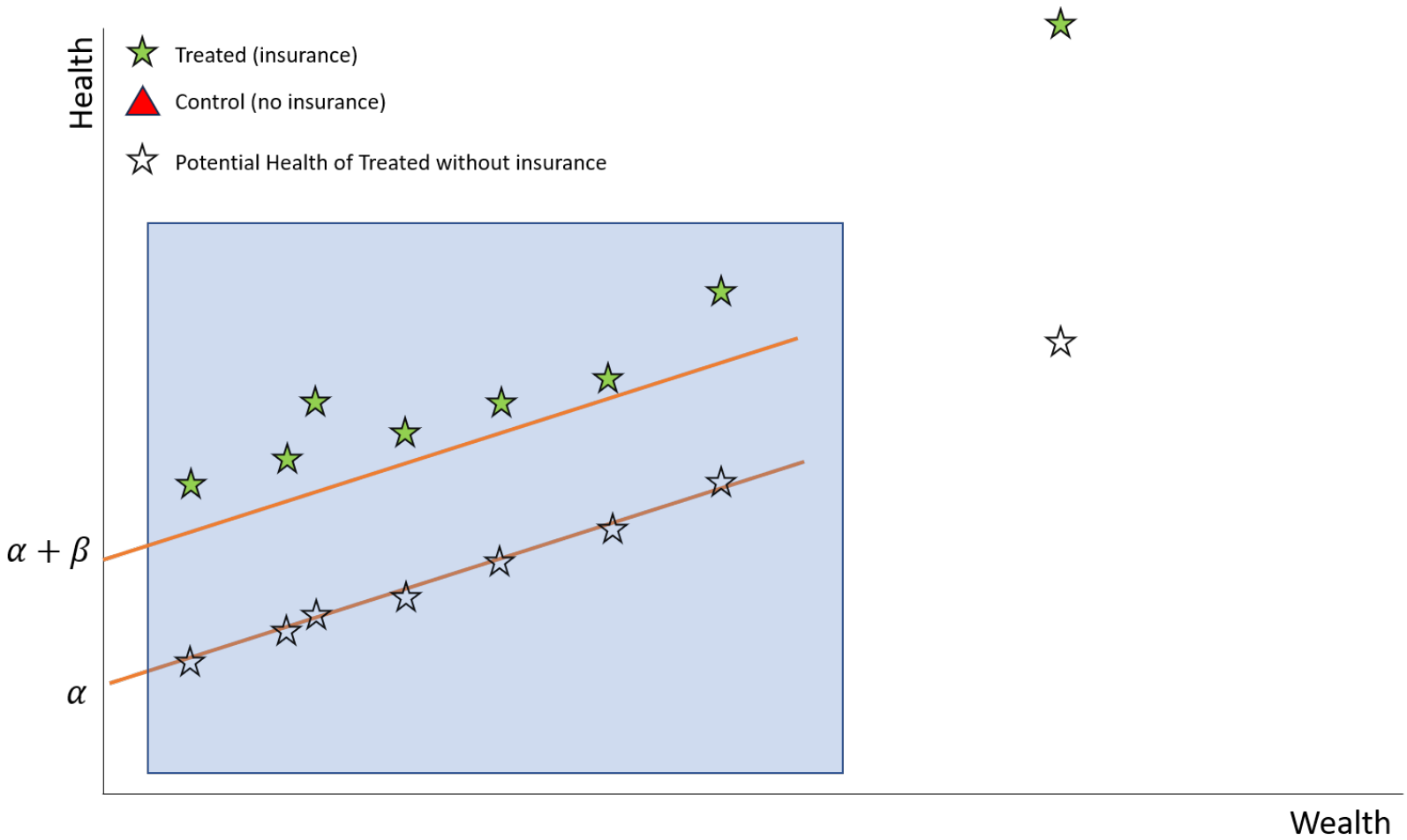
- There are no control units with such high wealth in our dataset
- We are extrapolating beyond on our common support

Q: So what's a solution?

A: RA + Matching

RA + Matching Solution

Run OLS only on the matched data



Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...

Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...
- IPTW may be biased and inefficient, and

Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...
- IPTW may be biased and inefficient, and
- RA may be biased and inefficient

Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...
- IPTW may be biased and inefficient, and
- RA may be biased and inefficient
- Meet the Doubly Robust Estimator (DR), which combines IPTW & RA

Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...
- IPTW may be biased and inefficient, and
- RA may be biased and inefficient
- Meet the Doubly Robust Estimator (DR), which combines IPTW & RA
- If either IPTW and/or RA is correctly specified, then DR is unbiased and might be more efficient

Doubly Robust TE Estimates

- If your model is misspecified (e.g. doesn't include all Good Controls)...
- IPTW may be biased and inefficient, and
- RA may be biased and inefficient
- Meet the Doubly Robust Estimator (DR), which combines IPTW & RA
- If either IPTW and/or RA is correctly specified, then DR is unbiased and might be more efficient
- And if both are wrong, DR may be less biased and more efficient than either IPTW or RA alone

Doubly Robust Estimator Combines IPTW and RA

Recall the IPTW estimator

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

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

Doubly Robust Estimator Combines IPTW and RA

Recall the IPTW estimator

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

where $\hat{e}_i$ is an estimate of the propensity score for unit i.
Lunceford and Davidian (2004) define

$$\begin{aligned} DR = &\frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i Y_i - (D_i - \hat{e}_i)\hat{E}[Y_i|D_i=1,X_i]}{\hat{e}_i}\right) \\ &- \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(1-D_i)Y_i + (D_i - \hat{e}_i)\hat{E}[Y_i|D_i=0,X_i]}{1-\hat{e}_i}\right) \end{aligned}$$

Steps for the RA-IPTW DR Estimator

1. Estimate the propensity scores $\hat{e}_i$
2. Estimate a regression model for the outcome only for the treated group $\hat{E}[Y_i|D_i = 1, X_i]$
3. Estimate the same regression model only for the control group $\hat{E}[Y_i|D_i = 0, X_i]$
4. Insert these components into the DR formula

Steps for the RA-IPTW DR Estimator

1. Estimate the propensity scores $\hat{e}_i$
2. Estimate a regression model for the outcome only for the treated group $\hat{E}[Y_i|D_i = 1, X_i]$
3. Estimate the same regression model only for the control group $\hat{E}[Y_i|D_i = 0, X_i]$
4. Insert these components into the DR formula

- Standard Errors can be computed analytically or via bootstrap
- Lunceford and Davidian (2004) and Emsley (2007) reveal controversy over choice of controls variates in the IPTW estimator

Outline

1. Case Study Overview
2. Matching as Linear Regression
3. Regression Adjustment (RA)
   - RA and Selection Bias
   - Regression + Matching
   - Treatment Effects in RA

RA Estimates the ATE

- ATE measures the treatment effect averaged across the entire sample

RA Estimates the ATE

- ATE measures the treatment effect averaged across the entire sample
- OLS reveals the marginal effect by fitting an MSE line through the averages of X and Y across the entire sample

RA Estimates the ATE

- ATE measures the treatment effect averaged across the entire sample
- OLS reveals the marginal effect by fitting an MSE line through the averages of X and Y across the entire sample
- By default, RA produces the ATE.

RA Estimates the ATE

- ATE measures the treatment effect averaged across the entire sample
- OLS reveals the marginal effect by fitting an MSE line through the averages of X and Y across the entire sample
- By default, RA produces the ATE.
- But we can extract the ATT and ATC with some care.

1. $Y = \alpha + \beta D + \gamma X + \delta D \times X + e \rightarrow \hat{\Theta} = \left\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right\}$

g-Computation

The algorithm

1. $Y = \alpha + \beta D + \gamma X + \delta D \times X + e \rightarrow \hat{\Theta} = \left\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right\}$
2. Switch all units as if they were treated. Predict outcome: $\tilde{Y}_{1i} = \left\{Y_{1i}(\hat{\Theta}, X_i) \text{ for } D_i = 0, Y_i \text{ for } D_i = 1\right\}$

g-Computation

The algorithm

1. $Y = \alpha + \beta D + \gamma X + \delta D \times X + e \rightarrow \hat{\Theta} = \left\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right\}$
2. Switch all units as if they were treated. Predict outcome:
   $\tilde{Y}_{1i} = \left\{ Y_{1i}(\hat{\Theta}, X_i) \text{ for } D_i = 0, Y_i \text{ for } D_i = 1 \right\}$
3. Switch all units as if they were untreated. Predict outcome:
   $\tilde{Y}_{0i} = \left\{ Y_{0i}(\hat{\Theta}, X_i) \text{ for } D_i = 1, Y_i \text{ for } D_i = 0 \right\}$

g-Computation

The algorithm

1. $Y = \alpha + \beta D + \gamma X + \delta D \times X + e \rightarrow \hat{\Theta} = \left\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right\}$
2. Switch all units as if they were treated. Predict outcome:
   $\tilde{Y}_{1i} = \left\{ Y_{1i}(\hat{\Theta}, X_i) \text{ for } D_i = 0, Y_i \text{ for } D_i = 1 \right\}$
3. Switch all units as if they were untreated. Predict outcome:
   $\tilde{Y}_{0i} = \left\{ Y_{0i}(\hat{\Theta}, X_i) \text{ for } D_i = 1, Y_i \text{ for } D_i = 0 \right\}$
4. Compute treatment effects (See next slide)

Treatment Effects via g-Computation

Constructing ATE/ATT/ATC

$$ATE = \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{1i} - \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{0i}$$

Treatment Effects via g-Computation

Constructing ATE/ATT/ATC

$$ATE = \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{1i} - \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{0i}$$

$$ATT = \frac{1}{N^{Treated}}\sum_{i=1}^{N^{Treated}}\tilde{Y}_{1i} - \frac{1}{N^{Treated}}\sum_{i=1}^{N^{Treated}}\tilde{Y}_{0i}$$

Treatment Effects via g-Computation

Constructing ATE/ATT/ATC

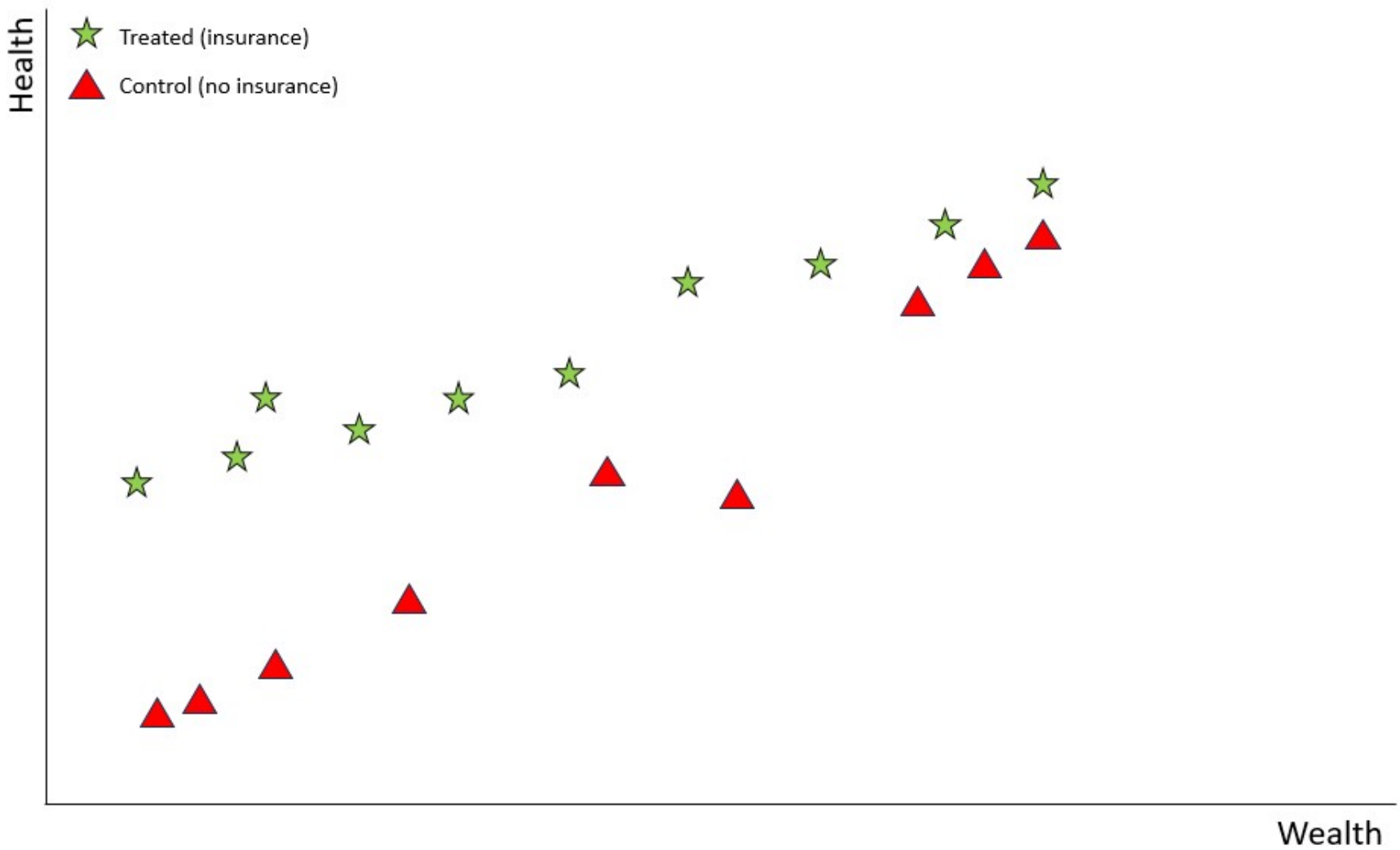
$$ATE = \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{1i} - \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_{0i}$$

$$ATT = \frac{1}{N^{Treated}}\sum_{i=1}^{N^{Treated}}\tilde{Y}_{1i} - \frac{1}{N^{Treated}}\sum_{i=1}^{N^{Treated}}\tilde{Y}_{0i}$$

$$ATC = \frac{1}{N^{Control}}\sum_{i=1}^{N^{Control}}\tilde{Y}_{1i} - \frac{1}{N^{Control}}\sum_{i=1}^{N^{Control}}\tilde{Y}_{0i}$$

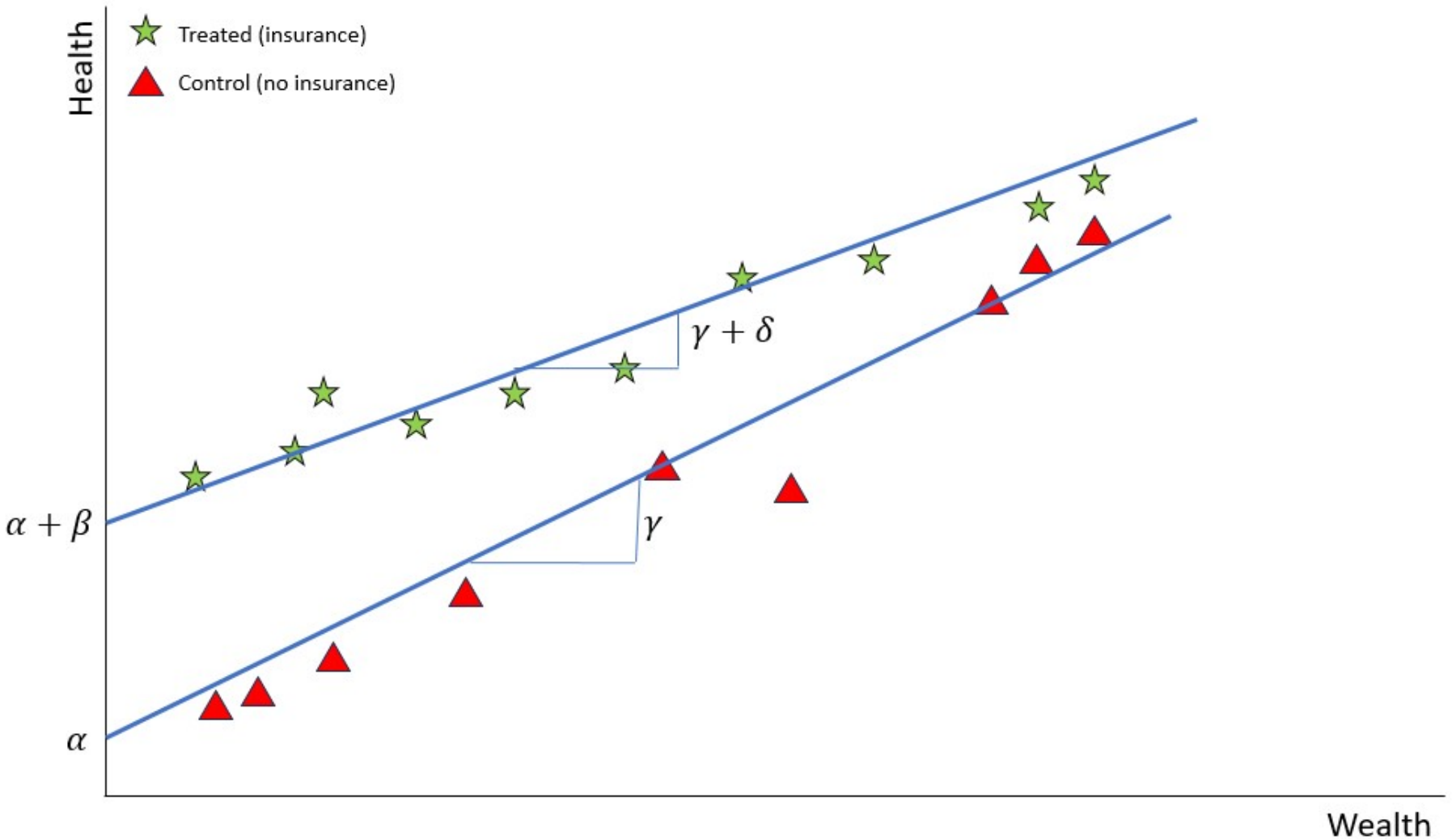
g-Computation Intuition for Treatment Effects

Given some data...



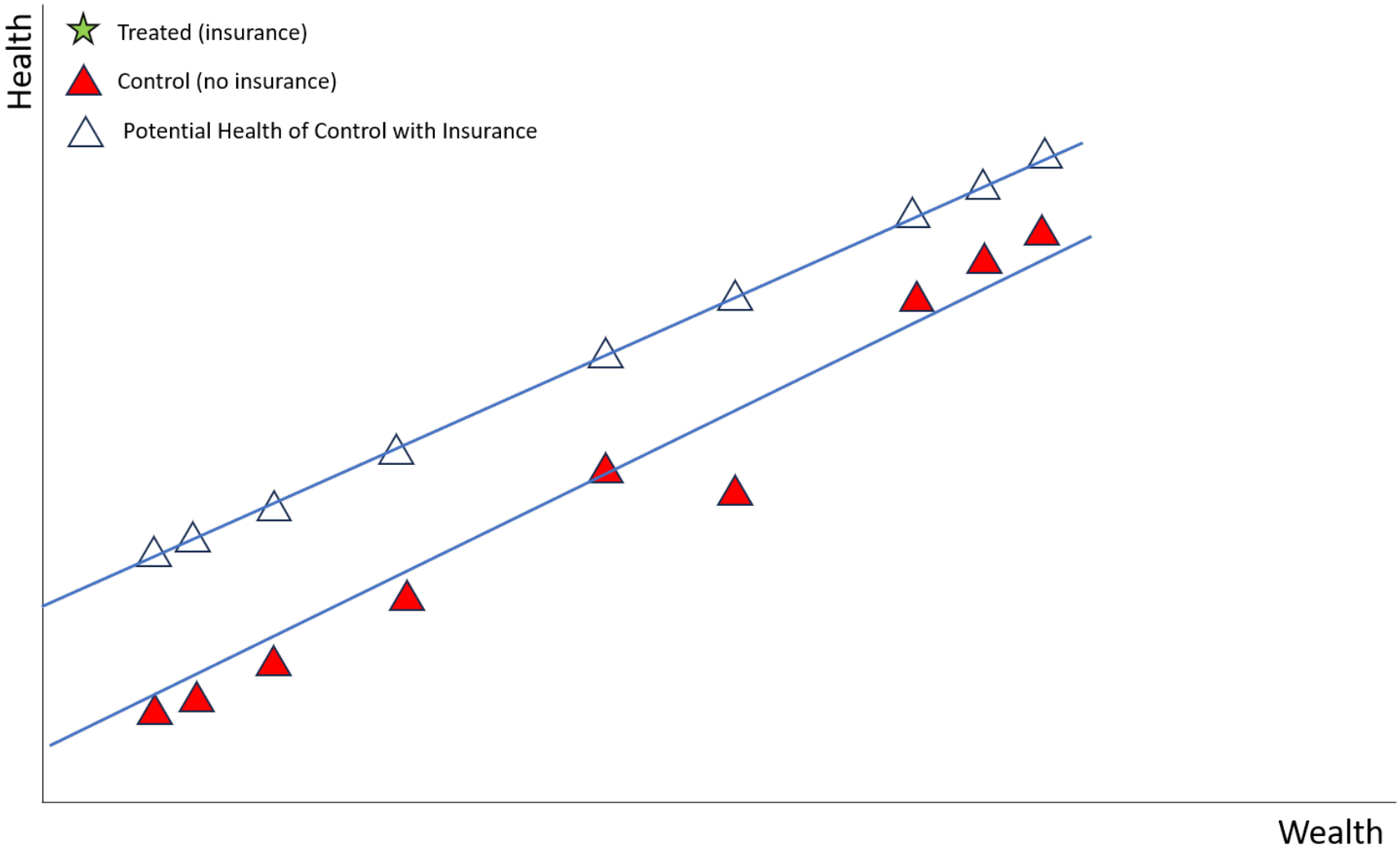
g-Computation Intuition for Treatment Effects

...we fit a regression line and grab parameter estimates $\hat{\Theta}$



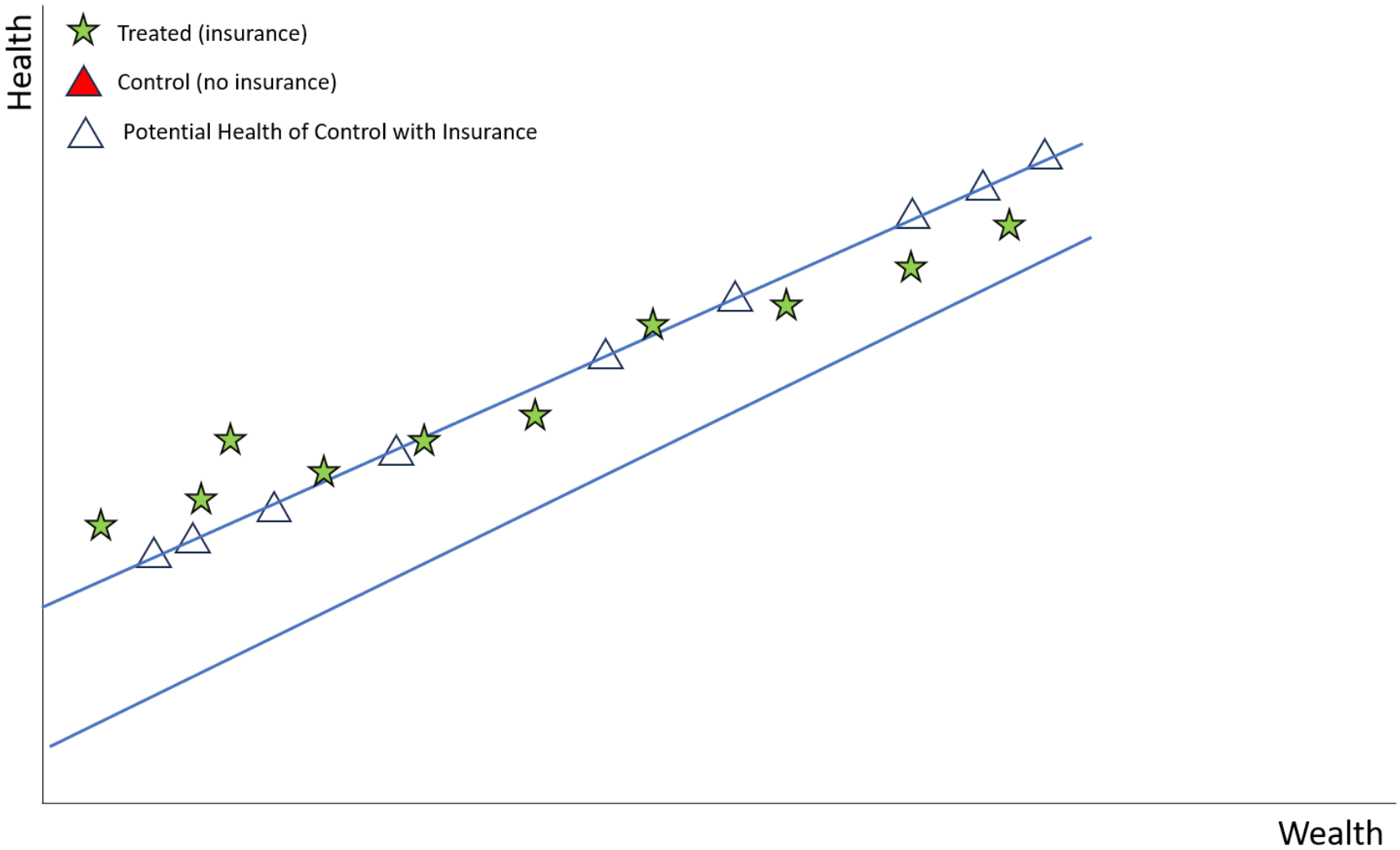
g-Computation Intuition for ATE

1) Use $\hat{\Theta}$ to estimate counterfactual for each control unit



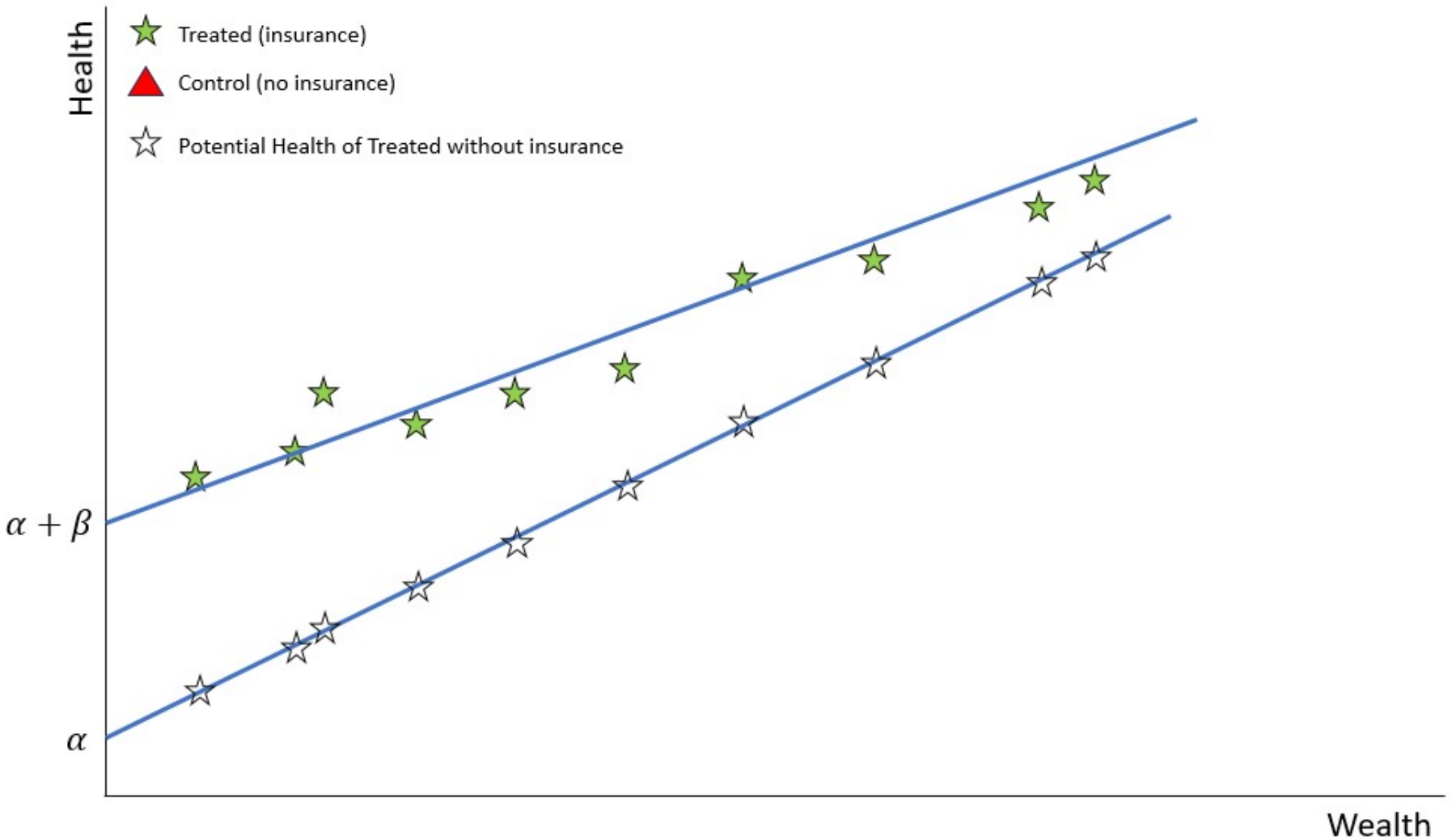
g-Computation Intuition for ATE

2) Combine with observed health of treated sample
3) Compute average of all green stars and blank triangles



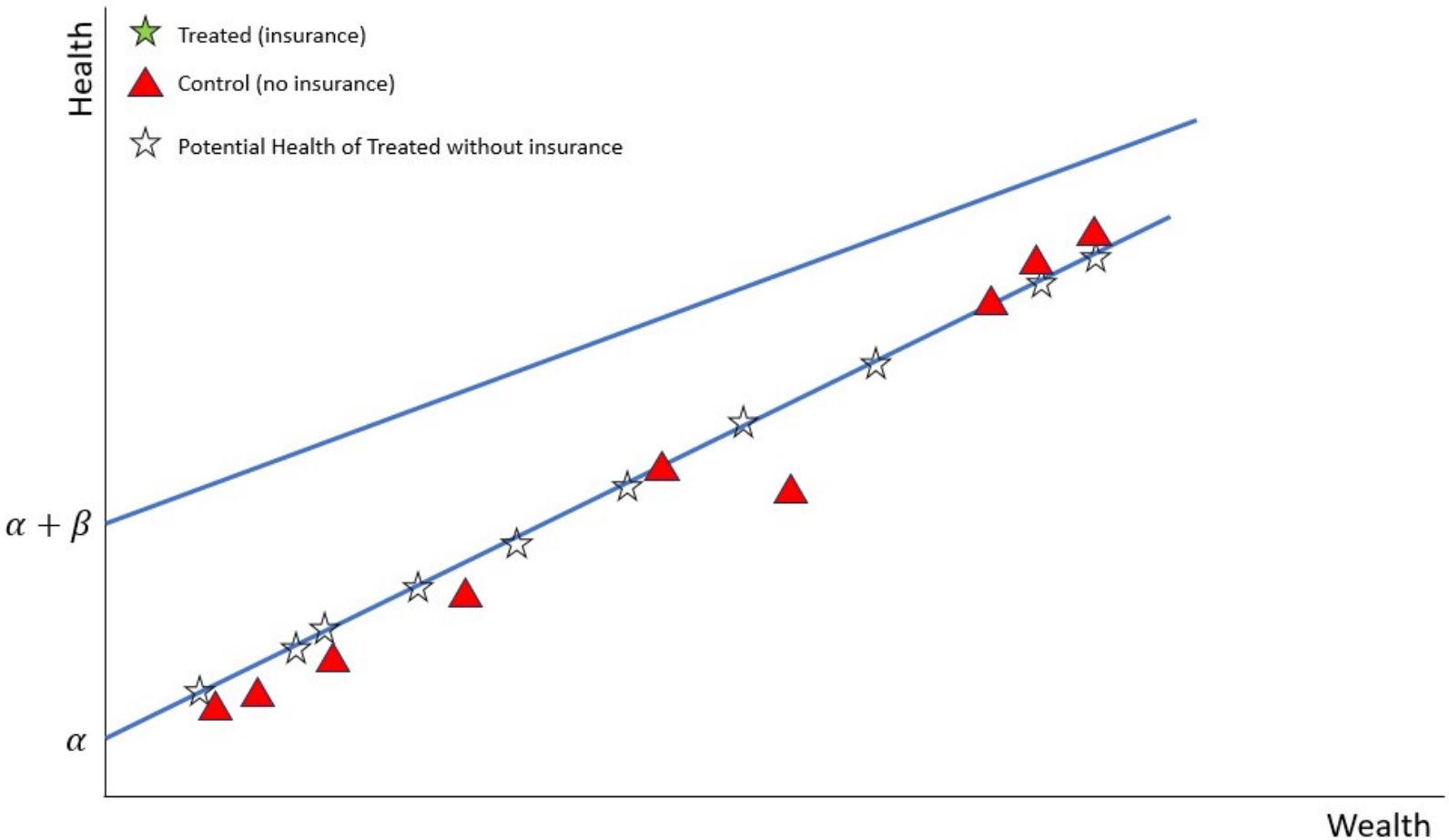
g-Computation Intuition for ATE

4) Use $\hat{\Theta}$ to estimate counterfactual for each treated unit



g-Computation Intuition for ATE

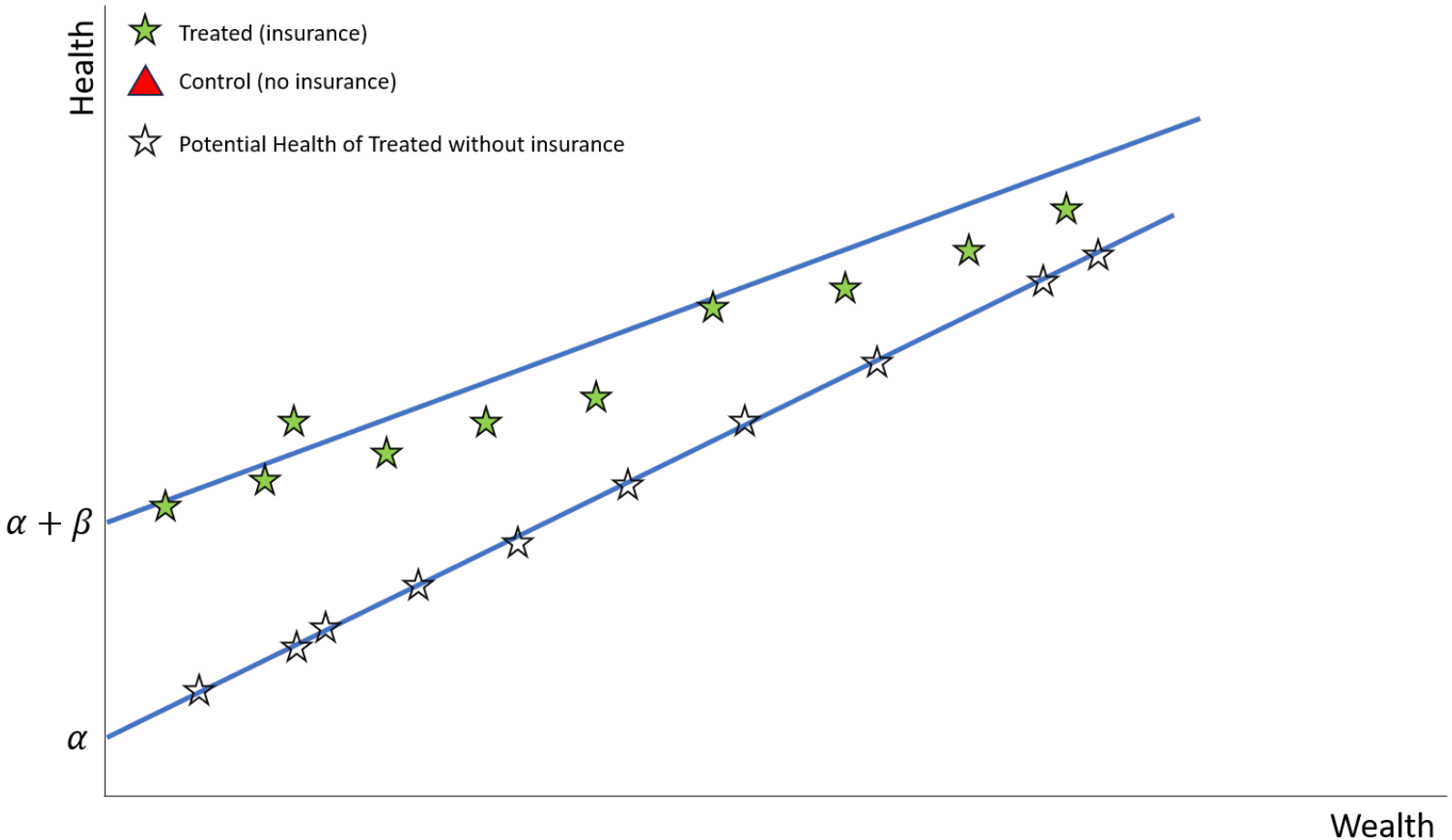
5) Combine with observed health of control sample
6) Compute average of all red triangles and blank stars
ATE = average from 3 minus average from 6



g-Computation Intuition - ATT

For ATT focus only on the treated units

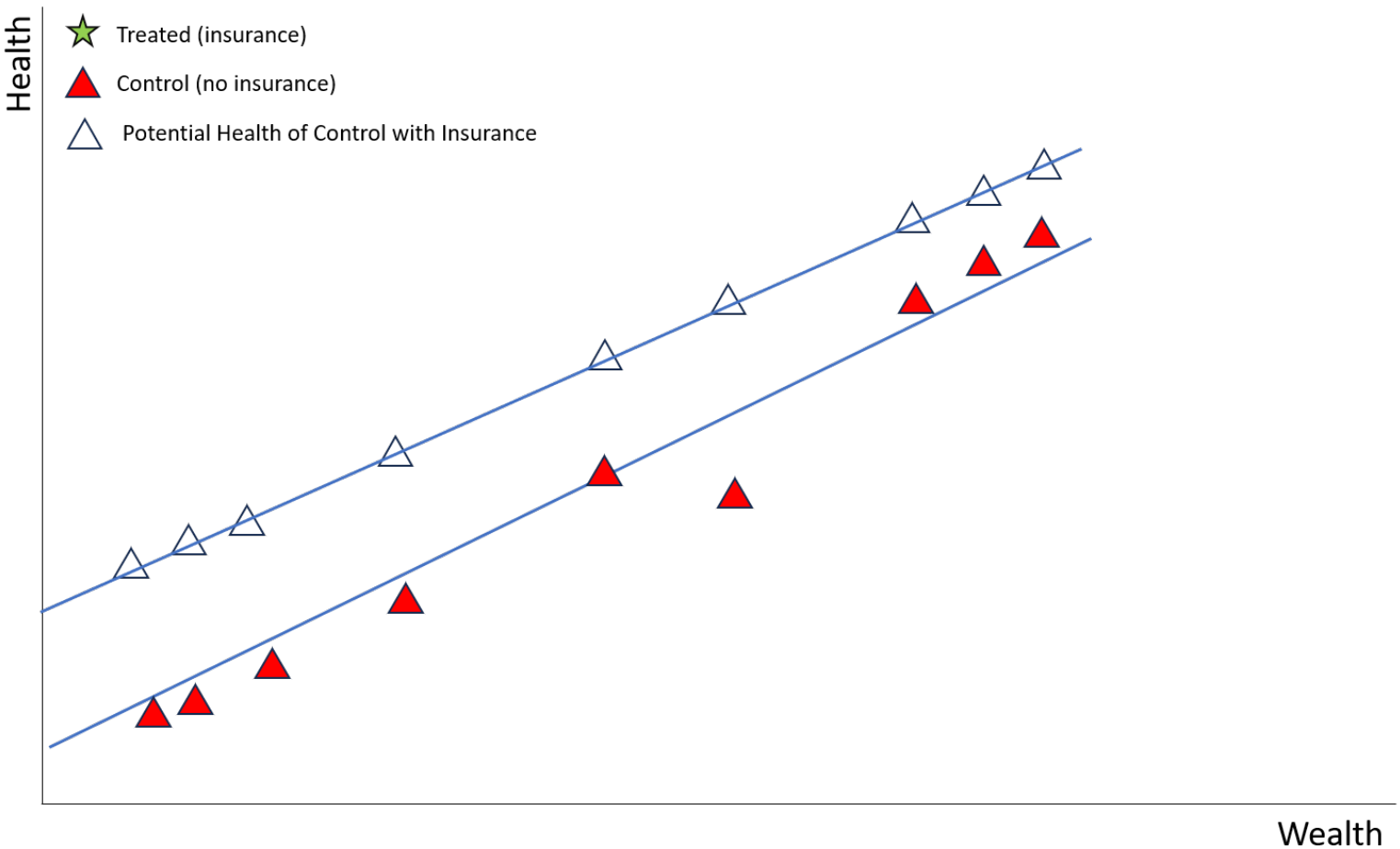
0) Use the same $\hat{\theta}$ estimates from ATE
1) Compute average of green stars
2) Compute average of blank stars
3) ATT = average from step 1 minus average from step 2



g-Computation Intuition - ATC

For ATC focus only on the control units

0) Use the same $\hat{\theta}$ estimates from ATE
1) Compute average of blank triangles
2) Compute average of red triangles
3) ATC = average from step 1 minus average from step 2



Regression Adjustment - Summary

- Regression Adjustment (RA) is the application of regression (OLS) to control for a confounder's impact on treatment effect
- OLS for prediction uses "X" to predict "Y"
- OLS for causality uses "X" to estimate $Y_1 - Y_0$
- Use care in choosing "Good" controls
- Combining Matching and Regression can be powerful