ECON 1123 Section 4

Slides at github.com/cjleggett/1123-section

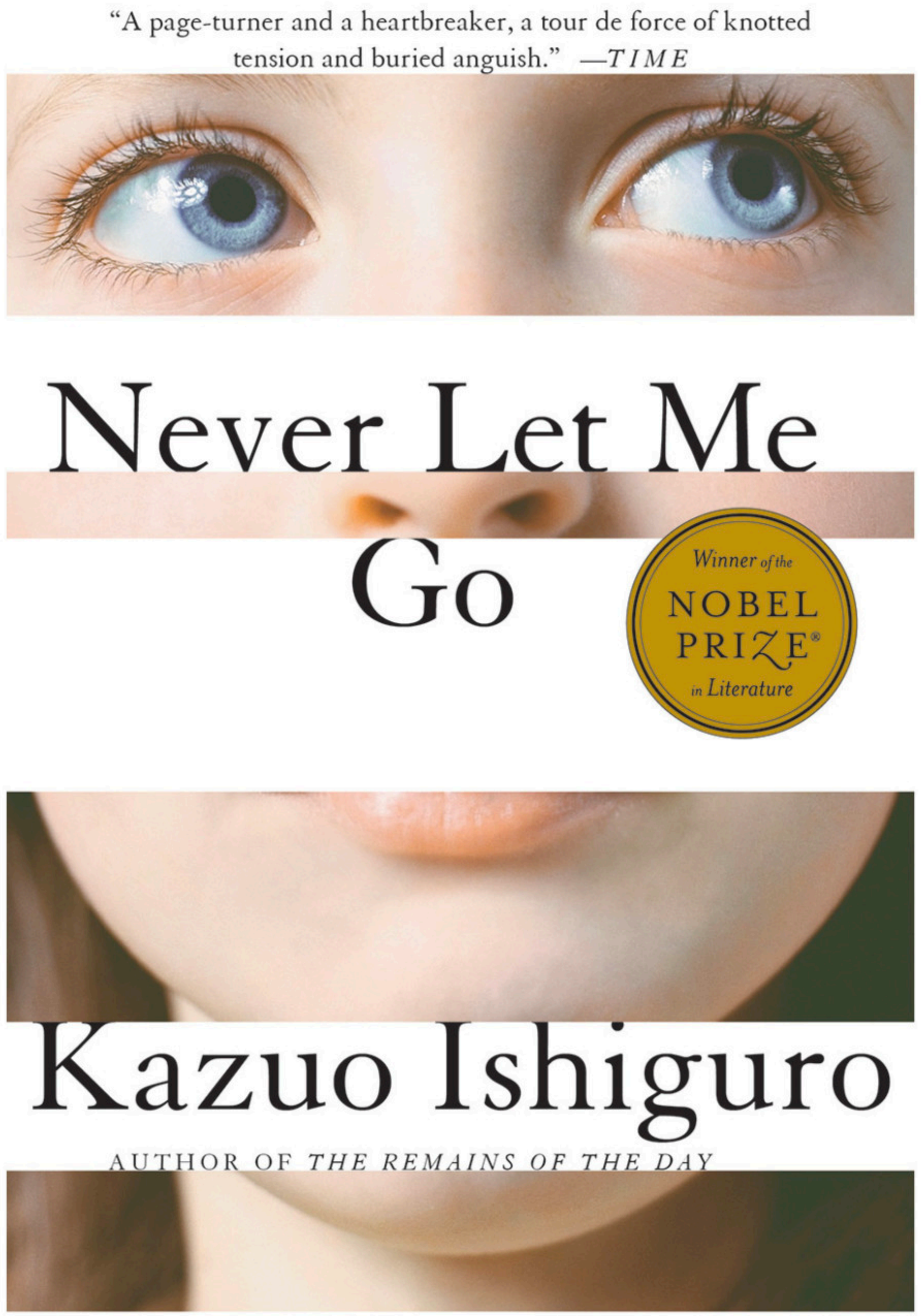
Outline

- Name Circle
- Pset Feedback
- Lecture Recap / Questions
- Examples + Practice

Name Circle

Name Circle

- Name
- Favorite Book (or a good book)

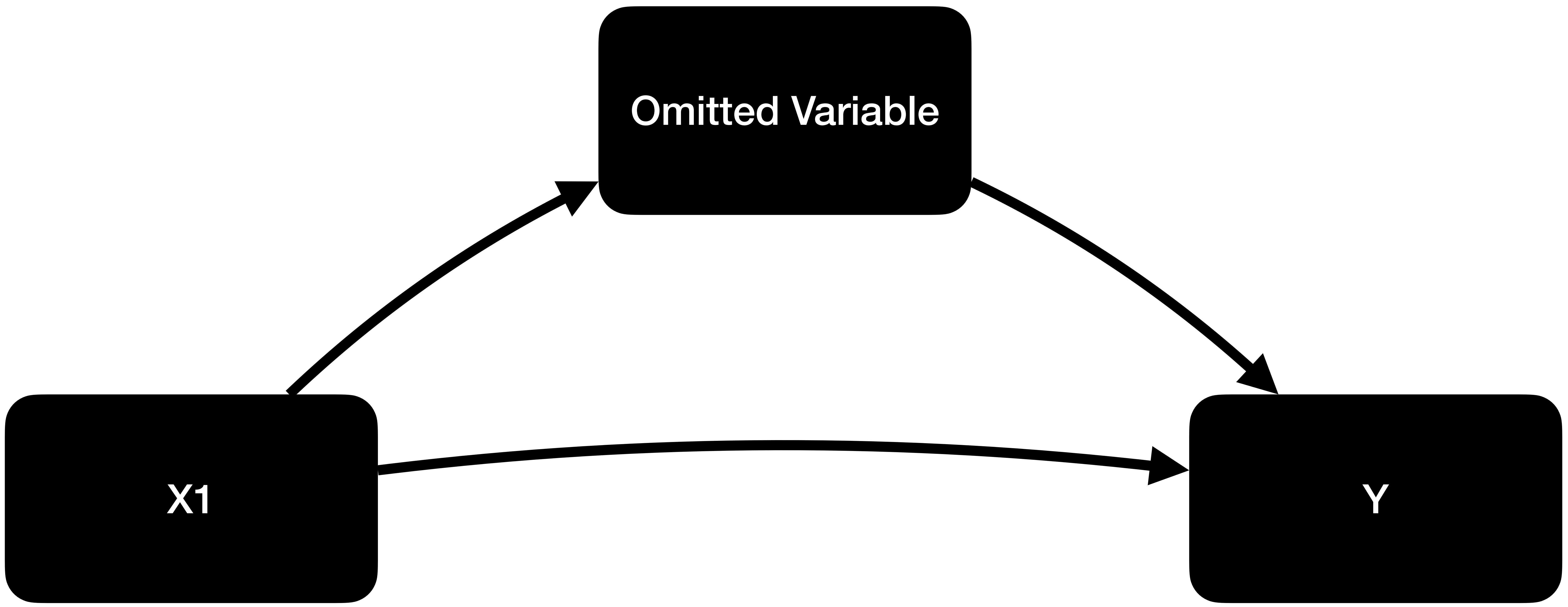


Problem Set Feedback

Great Work!

- You all did really well on a much more difficult pset!
- Just want to clear up OVB

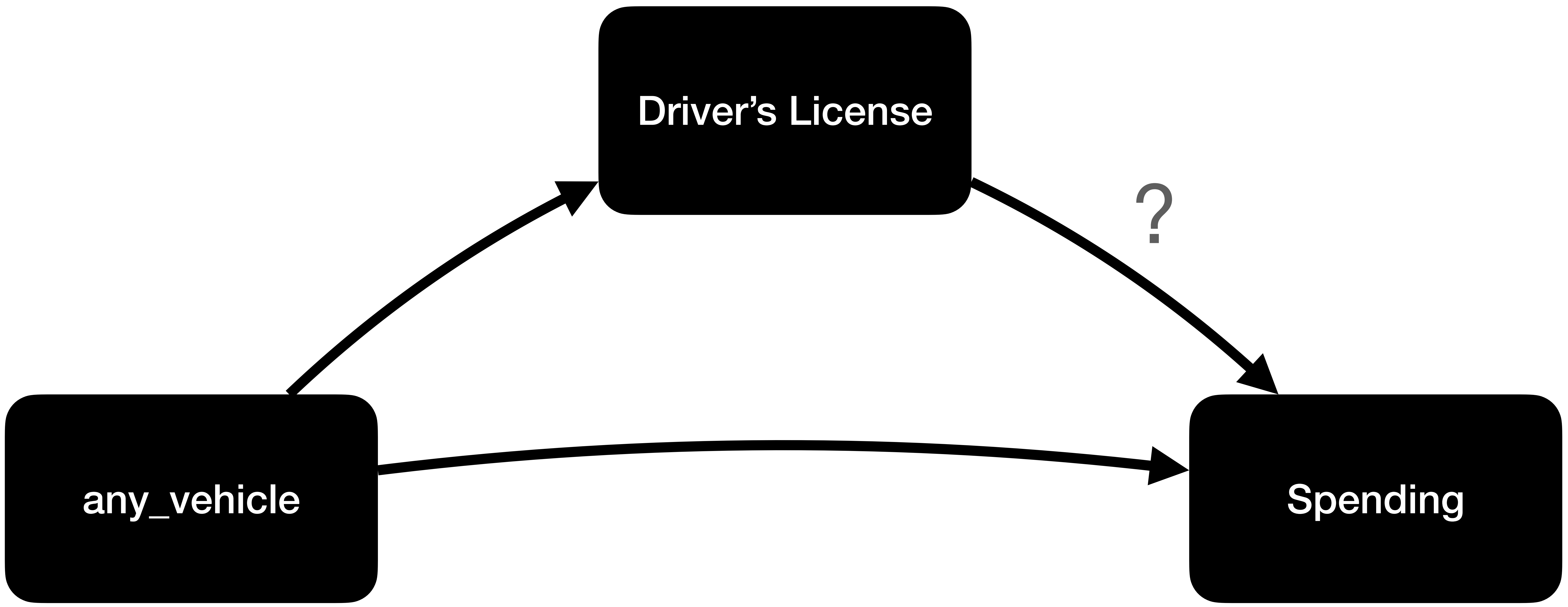
Choosing an Omitted Variable



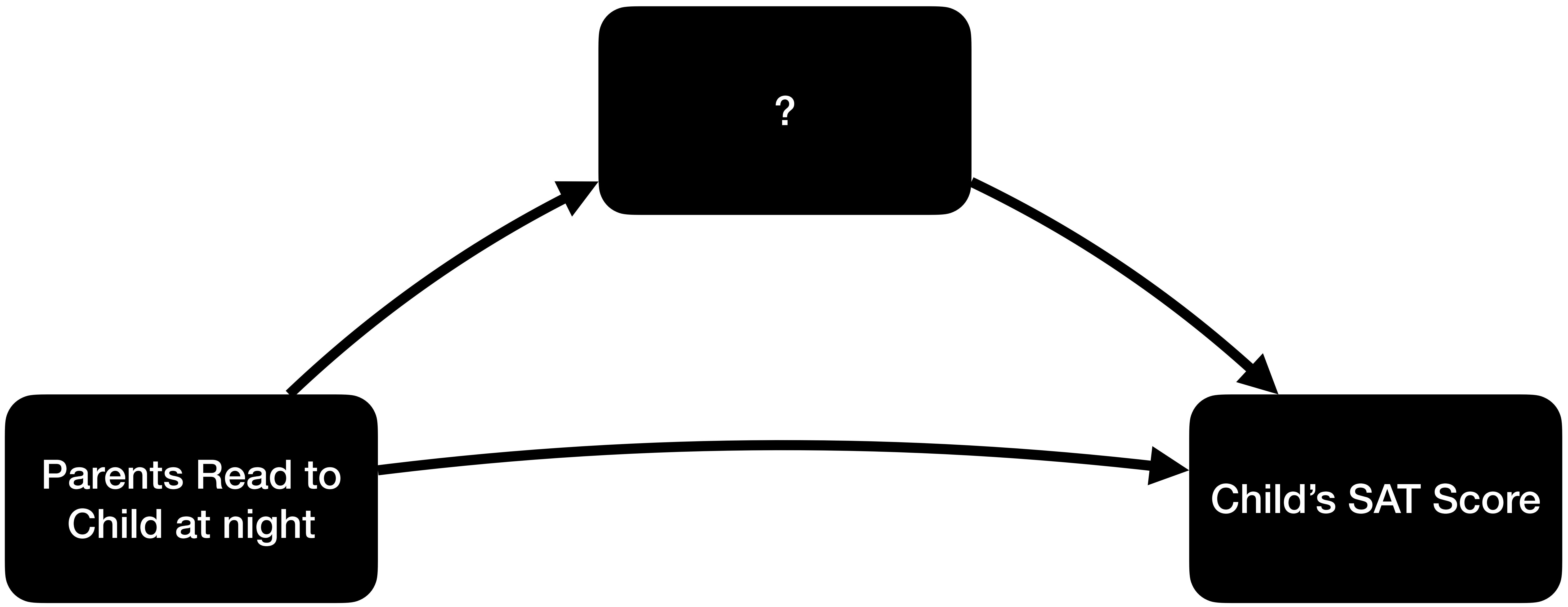
Choosing an Omitted Variable

Choosing an Omitted Variable

Choosing an Omitted Variable



OVB Practice!



Steps in OVB Problem

1. Choose an omitted variable

2. Sign β_2: correlation between omitted variable and outcome. Provide economic or real-world intuition for this sign.

3. Sign γ_1: correlation between omitted variable and variable of interest. Provide economic or real-world intuition for this sign.

4. Based on the above two, say the bias is positive or negative.

5. Conclude whether we have overstated or understated the causal effect. (Is α_1 too big or too small?)

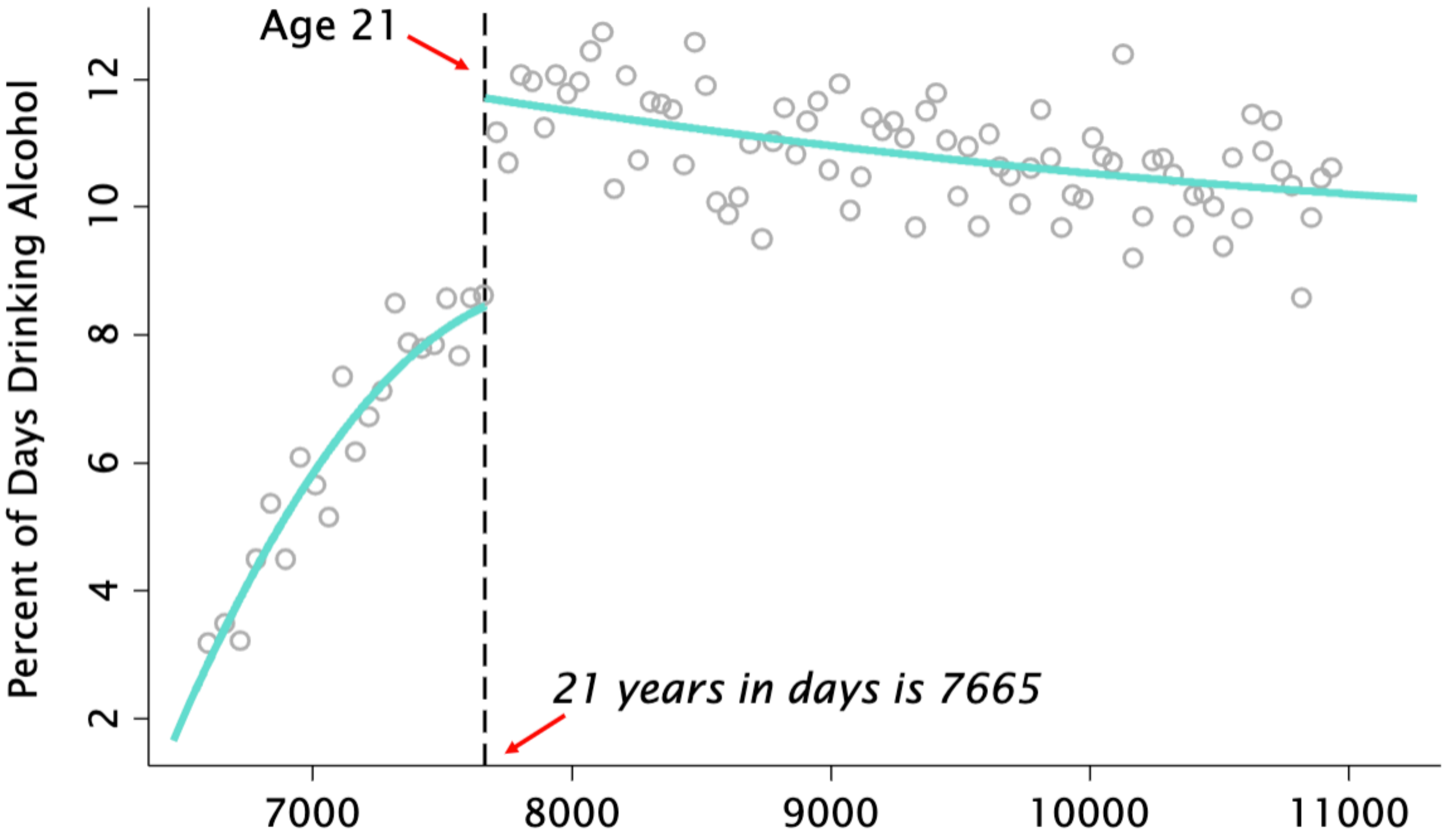
6. (Optional) Draw a number line.

Lecture Recap

Big Idea: Regression Discontinuity

- We can’t get true randomness, so sometimes we settle for almost randomness!
- Because of government/organization policies, we often have semi-arbitrary cutoff points (start drinking at 21, college has 3.25 GPA cutoff, National Merit Scholarship has PSAT cutoff)
- We often assume that people directly on either side of the cutoff are pretty similar, so we take the side of the cutoff as random!
- It’s important to check this assumption!

Graphing



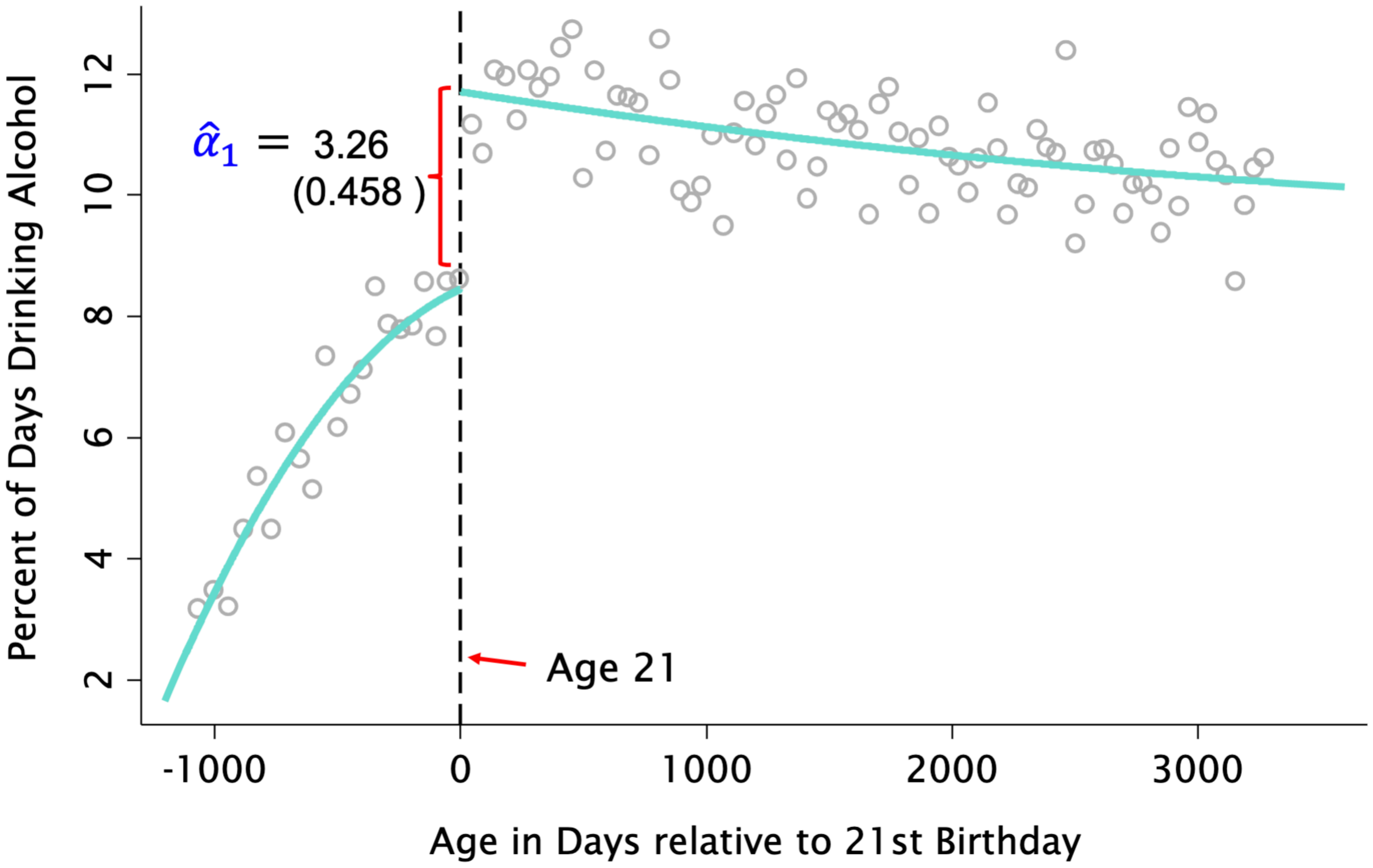
Age 21
21 years in days is 7665
Percent of Days Drinking Alcohol
2
4
6
8
10
12
7000
8000
9000
10000
11000

Included in Our Regression

- Indicator Variable for “Over Cutoff”
- “Running Variable” centered at cutoff, and exponents (makes math easier)
- Interactions between the above two
- Which Coefficient do we care about?

x

New Graph



$\hat{\alpha}_1$ = 3.26
(0.458)
Age 21
Percent of Days Drinking Alcohol
2
4
6
8
10
12
-1000
0
1000
2000
3000
Age in Days relative to 21st Birthday

Challenges

- Functional Form Bias: If it doesn't fit the data well, we may have functional form bias. We can check for this visually and by seeing how sensitive the coefficient we care about is to change in polynomial
- Misreporting: What if people misreport measurements for various reasons? Eg. Men's heights on dating apps (really!)

Linear Probability Model

- Just normal OLS
- Easier to interpret coefficients
- Doesn’t limit outcomes to between 0 and 1!

$$E[Y_i|X_i, W_i] = \Pr(Y_i = 1|X_i, W_i)$$
$$= \alpha_0 + \alpha_1 X_i + \alpha_2 W_i$$

Probit Model

- Wraps estimated coefficients in the CDF of the standard normal
- Values must be between 0 and 1
- Value within equation is a z-score

$$E[Y_i|X_i, W_i] = \Pr(Y_i = 1|X_i, W_i)$$

$$= \Phi(\beta_0 + \beta_1 X_i + \beta_2 W_i)$$

Logit Model

- Wraps estimated coefficients in the CDF of the logistic
- Values must be between 0 and 1
- Value within equation is a z-score

$$
\begin{aligned}
E[Y_i|X_i, W_i] &= \Pr(Y_i = 1|X_i, W_i) \\
&= F(\gamma_0 + \gamma_1 X_i + \gamma_2 W_i) \\
&= \frac{1}{1 + \exp\{-(\gamma_0 + \gamma_1 X_i + \gamma_2 W_i)\}} \\
&= \frac{\exp\{\gamma_0 + \gamma_1 X_i + \gamma_2 W_i\}}{1 + \exp\{\gamma_0 + \gamma_1 X_i + \gamma_2 W_i\}}
\end{aligned}
$$

Predicted Change in Probability

- Can’t always predict 1 or 0 for everyone, so we use models to predict probabilities between 0 and 1
- Can’t always interpret the coefficients well, so we predict the **change in probability** at a point (kind of like how we interpreted predicted change in polynomial models)
- This is not too hard: just plug in numbers and subtract
- Standard errors of these predictions are difficult! We use something called the delta method that we don’t need to worry about

Which one to use?

- In most of our cases, we care about causal inference, not predictions
- So it typically doesn’t matter which one we use!
- Except linear coefficients are easier to interpret…

Exercises!