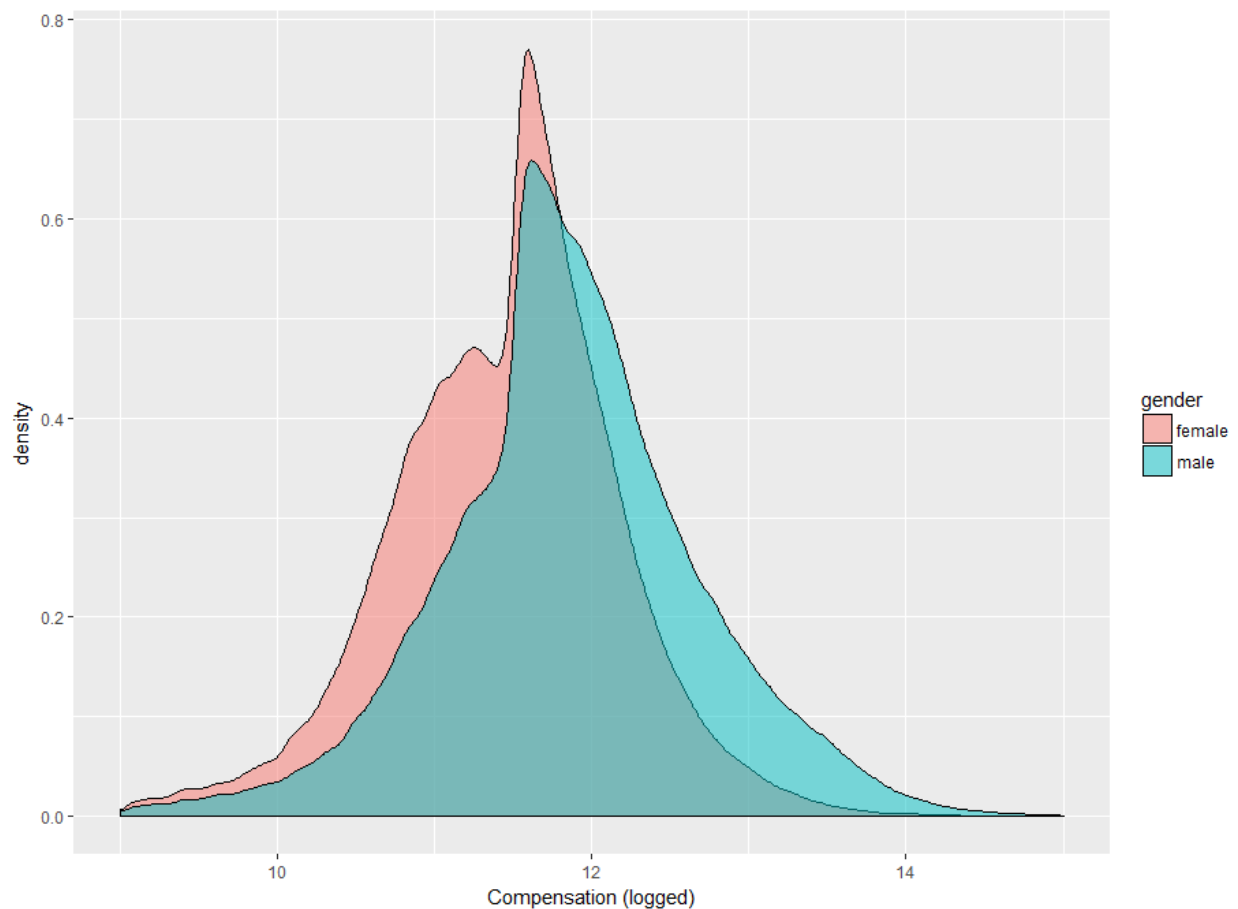


CPP 524: PE II Research Design

Prof. Jesse Lecy

Advanced Modeling: Part I

For this part of the final assignment you will be using a dataset that examines compensation of nonprofit executive directors from the years 2012-2013. The data is extracted from the IRS E-Filer database available on AWS. A gender variable has been added by matching the first names of the executive director to Census birth certificate data, and predicting gender based upon the typical gender of the individuals with that name (see the PROPORTION_FEMALE variable). We are interested in discrepancy of salaries of the executive directors by gender.



Codebook:

FILEREIN	– Tax ID of the nonprofit
TAXYR	– Year of the tax record (this data is from 2012 and 2013)
FILERNAME1	– Name of the nonprofit
STATE	– Location of the nonprofit
RULEDATE	– Year and month the nonprofit was granted status
NPAGE	– Nonprofit age
REVENUE	– Total annual revenue for the nonprofit
ASSETS	– Total assets of the nonprofit
PERSONNM	– Name of the Executive Director
TITLETXT	– Title of the Executive Director
AVGHRS	– Average hours worked each week
SALARY	– Annual salary for the Executive Director
GENDER	– Typical gender for someone with that first name
PROPORTION_FEMALE	– The proportion of babies born with that first name that are female
M2012CEO	– Was there a male executive director in 2012? 1=yes, 0=no
TREAT	– Did the organization hire a new CEO in 2013 with a different gender?
POST	– Dummy variable for the second year: 1=2013, 0=2012
NTMAJ12	– Subsector of the nonprofit
	<ul style="list-style-type: none">• AR Arts, culture, and humanities• BH Education, higher• ED Education• EH Hospitals• EN Environment• HE Health• HU Human services• IN International• MU Mutual benefit• PU Public and societal benefit• RE Religion• UN Unknown

Part I: The Naïve Model

- (A) Test the hypothesis that men get paid more than women using only the salary and gender variables. Run the following regression:

$$\text{Pay} = b_0 + b_1 * \text{male}$$

What is the difference in pay? Is the relationship statistically significant? Interpret your results.

- (B) What sort of research design is this? Reflexive, post-test only, or pre-post with comparison (diff-in-diff)?
- (C) If you define pay discrimination as paying two people different amounts for the exact same work, do you feel this is a good estimate of pay discrimination? Why not? What are potential problems with the naïve model?

Part II: Improved Cross-Section Model

- (A) Improve upon your model from Part I by adding control variables to account for omitted variables. Interpret your results.
- (B) How did your results change, if any?

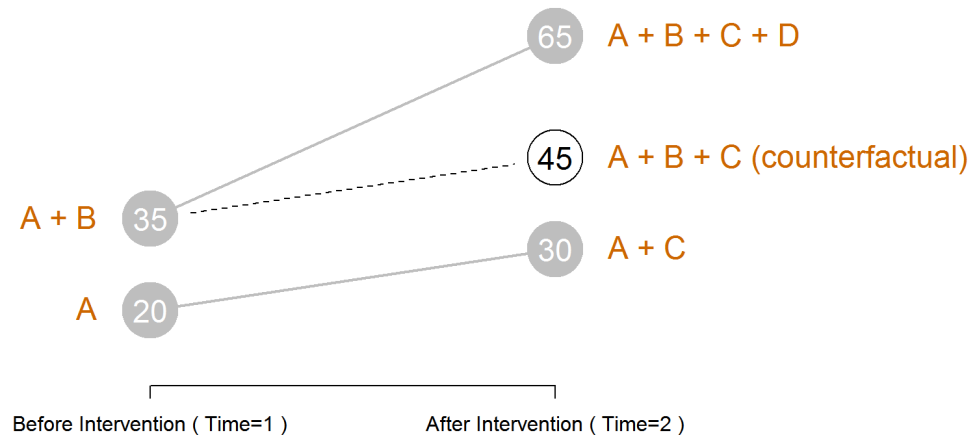
Part III: The Difference-In-Difference Approach

Instead of adding more control variables to your regression to account for omitted variable bias, you can also change your estimator. Recall that each has different assumptions or requirements to be unbiased. We worry about the naïve model above because it is not clear that the assumptions are met.

- (A) Use a difference-in-difference model to examine how the pay structure changes when a nonprofit has a male executive director and hires a new female Executive Director (ED). The “treatment” in this scenario is a change in the gender of the ED.

The difference-in-difference model:

$$Y = B_0 + B_1 \cdot \text{Treat} + B_2 \cdot \text{Post} + B_3 \cdot \text{Treat} \cdot \text{Post} + e$$



$$Y = A + B \cdot \text{Treat} + C \cdot \text{Post} + D \cdot \text{Treat} \cdot \text{Post} + e$$

We must first split the data into two separate datasets using the variable M2012CEO. This variable is coded as a 1 if the Executive Director is a male in 2012, and 0 if it is a female. To run the difference-in-difference model we want to focus on cases where the gender of the CEO changes.

```
dat$TREAT_POST <- dat$TREAT * dat$POST
dat.male <- dat[ dat$M2012CEO == 1 , ]
m1 <- lm( SALARY ~ TREAT + POST + TREAT_POST, data=dat.male )
stargazer( m1, type="html", digits=0,
omit.stat = c("rsq","f","ser","adj.rsq"),
column.labels = c("TREATMENT: Male to Female CEO"),
intercept.bottom = FALSE,
covariate.labels = c("CONSTANT (A)", "TREATMENT (B)",
"POST (C)", "TREATMENT-POST (D)" ) )
```

- (B) What is the counterfactual in this case? Is it a better estimate of the pay gap than cross-sectional, observational models from Part I and Part II?
- (C) Interpret the results of the model by answering the following questions:
 - (1) Are the nonprofits that switch from male to female EDs any different than those that hire another male ED?
 - (2) Do we see any secular trends (change in pay when a male replaces a male, i.e. the 'control' group)?
 - (3) Do we observe a (statistically-significant) pay gap for female EDs? How much is their pay different than what was expected?
- (D) What is the female ED counterfactual salary? What they would get paid if they were paid the same as male EDs?