

All Together

Part I

Fernando Hoces la Guardia
08/8/2022

Today's and Tomorrow's Lecture

- Review most of our tools for causal inference applying them to one policy issue:
 - The effect of education on earnings

We will do this in three steps:

- First we will use selection bias, potential outcomes, RCTs and regression to frame this causal question.
- Then we will learn our last important concept: bad controls.
- Finally we will see how we can use IV, DD and RDD to answers this question.

Earnings and Schooling

- Understanding what is behind a persons' earnings is a core question in economics:
 - From a economic growth perspective, its important to understand what drives growth in individual incomes.
 - From as distributional perspective, its important to understand why some earn so much more than others.
- One possible channel that could help us understand what causes income (growth and inequality) is education. Let's go ahead and define our causal question:
 - What is the effect of one additional year of schooling (S_i) on earnings (Y_i) in the US?
 - First approach: RCT as a thought exercise, compare with a simple difference in groups.

Definitions

- Outcome (Y_i): annual earnings
- Treatment (S_i): years of education
 - Important underlying assumption: no differences in education quality (i.e. one additional year of education in Cal is the same as one additional year of education in Trump University).
 - This is strong and unrealistic assumption, but provides a good starting point, and the methods here can (and have been) extended to account for differences in quality.
- Populations: typically a representative sample of US adults (but will vary across studies).

Answering the Causal Question: Before Expectations

- Simple difference in groups between individuals with one year difference in education. For example between 13 and 12 years of education:

Simple Difference in Groups =

[Average earnings among those with 13 years of schooling] –
[Average earnings among those with 12 years of schooling]

Answering the Causal Question: Before Expectations

- Simple difference in groups between individuals with one year difference in education. For example between 13 and 12 years of education:

Simple Difference in Groups =

[Average earnings among those with 13 years of schooling] –
[Average earnings among those with 12 years of schooling]

Simple Difference in Groups = $\text{Avg}[Earnings_i | Schooling_i = 13] - \text{Avg}[Earnings_i | Schooling_i = 12]$

Answering the Causal Question: Before Expectations

- Simple difference in groups between individuals with one year difference in education. For example between 13 and 12 years of education:

Simple Difference in Groups =

[Average earnings among those with 13 years of schooling] –
[Average earnings among those with 12 years of schooling]

$$\begin{aligned}\text{Simple Difference in Groups} &= \text{Avg}[Earnings}_i | \text{Schooling}_i = 13] - \\ &\quad \text{Avg}[Earnings}_i | \text{Schooling}_i = 12] \\ &= \text{Avg}[Y_i | S_i = 13] - \text{Avg}[Y_i | S_i = 12] \\ &= [\bar{Y} | S = 13] - [\bar{Y} | S = 12]\end{aligned}$$

Answering the Causal Question: Before Expectations

- Simple difference in groups between individuals with one year difference in education. For example between 13 and 12 years of education:

Simple Difference in Groups =

[Average earnings among those with 13 years of schooling] –
[Average earnings among those with 12 years of schooling]

$$\begin{aligned}\text{Simple Difference in Groups} &= \text{Avg}[Earnings}_i | \text{Schooling}_i = 13] - \\ &\quad \text{Avg}[Earnings}_i | \text{Schooling}_i = 12] \\ &= \text{Avg}[Y_i | S_i = 13] - \text{Avg}[Y_i | S_i = 12] \\ &= [\bar{Y} | S = 13] - [\bar{Y} | S = 12]\end{aligned}$$

- But what is the difference for many samples over the long run?

Expectations, Selection Bias & Potential Outcomes

- Expected value of simple difference in groups between individuals with one year difference in education (now for any given one year difference, $s + 1$ and s):

$$\mathbb{E}(\text{Simple Difference in Groups}) = \mathbb{E}(Y_i | S_i = s + 1) - \mathbb{E}(Y_i | S_i = s)$$

Expectations, Selection Bias & Potential Outcomes

- Expected value of simple difference in groups between individuals with one year difference in education (now for any given one year difference, $s + 1$ and s):

$$\begin{aligned}\mathbb{E}(\text{Simple Difference in Groups}) &= \mathbb{E}(Y_i | S_i = s + 1) - \mathbb{E}(Y_i | S_i = s) \\ &= \kappa + \underbrace{\mathbb{E}(Y_{i0} | S_i = s + 1) - \mathbb{E}(Y_{i0} | S_i = s)}_{\text{Selection bias}}\end{aligned}$$

- Do we think that the earnings of individuals who get one more year of education **would have been** similar to those that didn't get that additional year?
- If the answer is no, then we have selection bias.

Regression and CEF

- We can also rewrite the expected SDG as using regression (exactly if CEF is linear, as with a binary regressors, or with the best approximation if CEF is not linear):

$$Y_i = \alpha + \underbrace{\mathbb{E}(\text{Simple Difference in Groups})}_{\rho} \times S_i + e_i$$

- Now we have our first regression equation for our causal question of interest:

$$Y_i = \alpha + \rho S_i + e_i$$

- One advantage of having moved from SDG to regression is that we can try to control for other things that might be behind this selection bias.
- And to interpret ρ as percentage change (or "return") in earnings (from a 1-unit change in schooling), we replace the dependent variable with its logs ($\ln Y_i$)

Regression as Matching and OVB

- For example, maybe excluding experience (X_i) is creating some OVB. If the regression above is the short equation ($\ln Y_i = \alpha^s + \rho^s S_i + e_i^s$), what would be the long equation, auxiliary and OVB?

Regression as Matching and OVB

- For example, maybe excluding experience (X_i) is creating some OVB. If the regression above is the short equation ($\ln Y_i = \alpha^s + \rho^s S_i + e_i^s$), what would be the long equation, auxiliary and OVB?

$$\text{long: } \ln Y_i = \alpha^l + \rho^l S_i + \beta_1 X_i + e_i^l$$

$$\text{aux: } X_i = \pi_0 + \pi_1 S_i + u_i$$

$$\text{OVB: } \rho^s - \rho^l = \pi_1 \beta_1$$

- Given that schooling requires being out of the labor force, we should expect a negative correlation between X_i and S_i , hence $\pi_1 < 0$.
- Given that on-the-job skills improve with experience we should expect a positive relationship between $\ln Y_i$ and X_i (up to a point), hence $\beta_1 > 0$. This implies that $\pi_1 \beta_1 < 0$ and $\rho^s - \rho^l < 0$, or that $\rho^s < \rho^l$. Given that $\rho^s > 0$, this means that the short regression is underestimating the causal effect.

Results

- First estimates of this kind were produced by Jacob Mincer for a sample of 31,000 men in the US in the 1960s:

$$\ln Y_i = \alpha + 0.070S_i + e_i$$

(0.002)

$$\ln Y_i = \alpha + 0.107S_i + 0.081X_i - 0.0012X_i^2 + e_i$$

(0.002) (0.001) (0.00002)

- It is conventional to add the quadratic term for experience to capture the fact that experience increases earnings up to a point (about 30 years of experience).
- This result shows that, **if** we believe that controlling for years of experience is enough to remove OVB, each additional year of education **causes** an increase of almost 11% (10.7%) in earnings on average.

Other Possible OVBs

- The story that we are controlling for all possible factors that could cause OVB, is hard to believe.
- In economics a common catch all term to refer to commonly omitted factor is the variable "Ability".
- However, it is important to remember that not all omitted factors are "skills or talents" that are immutable, and that make the treatment group "better" than the control. For example other factors like, connections, ease which individuals conduct themselves (due to gender, race or class), or other factor could equivalently generate OVB here. We could hence include a second catch all variable and call it "Privileged".
- But to keep things simple will keep just one variable to model the omitted factor and refer to it a OF_i for other factors (corresponds to the variable A_i in the book).

OVB Analysis

For the case of "ability":

- High ability people are likely to, on average, find schooling less difficult, and get more of it (on average). Hence, $\delta_{OF,S} > 0$.
- High ability people are likely have, on average, higher wages. Hence $\gamma > 0$

For the case of "privileged":

- High privileged people are likely to, on average, find schooling less difficult, and get more of it (on average). Hence, $\delta_{OF,S} > 0$.
- High privileged people are likely have, on average, higher wages. Hence $\gamma > 0$

- For simplicity we go back to the initial short regression (without experience):

$$\text{short: } \ln Y_i = \alpha^s + \rho^s S_i + e_i^s$$

$$\text{long: } \ln Y_i = \alpha^l + \rho^l S_i + \gamma OF_i + e_i^l$$

$$\text{aux: } OF_i = \pi_0 + \delta_{OF,S} S_i + u_i$$

$$\text{OVB: } \rho^s - \rho^l = \underbrace{\delta_{OF,S} \gamma}_{\text{ability/privileged bias}}$$

- $OVB > 0$ for both cases!

Time to Bring More Controls?

- Why not just add more controls to the regression above?
- How about occupation?
 - Occupations is likely to matter for earnings, and schooling correlates with occupations.
 - Omitting it would lead to OVB?
- Everything that we have learned until now would point in the direction of "yes". But the answer is actually "no", and to understand why we need to explore the last concept of this course: bad controls.

Bad Controls

- A control should not be included (is bad) when it is, at least in part, **caused** by the treatment.
- Including a bad control leads to selection bias due to composition effects.
- To see why let's use the following simplified example:
 - Two occupations: blue collar (BC) and white collars (WC) and three type of individuals:
 1. a first minority of individuals for whom education has no effect: their earnings don't change (\$1000) and their occupation is BC with and without schooling.
 2. a majority of individuals for whom education **has an effect: their earnings go from 0 to \$2000**, and their occupation goes from BC to WC .
 3. a second minority of individuals for whom education also has no effect: their earnings don't change (\$4000) and their occupation is WC with and without schooling.

Toy Data for Bad Controls 1/2

- In this setting we know the potential outcomes of all individuals, and we randomly assign some to a treatment of one additional year of schooling and some to no additional schooling. Assuming a population of 10 we could have the following:

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w	TE_i	D	OC_i	Y_i^w
1	BC	BC	1	1	0	0	BC	1
2	BC	BC	1	1	0	1	BC	1
3	BC	WC	0	2	2	0	BC	0

Showing 1 to 3 of 10 entries

Previous

1

2

3

4

Next

Toy Data for Bad Controls 2/2

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w	TE_i	D	OC_i	Y_i^w
1	BC	BC	1	1	0	0	BC	1
2	BC	BC	1	1	0	1	BC	1
3	BC	WC	0	2	2	0	BC	0
4	BC	WC	0	2	2	1	WC	2
5	BC	WC	0	2	2	0	BC	0
6	BC	WC	0	2	2	1	WC	2
7	BC	WC	0	2	2	0	BC	0
8	BC	WC	0	2	2	1	WC	2
9	WC	WC	4	4	0	0	WC	4

Estimating Average Treatment Effects Without Controls

- Average wage for treatment group: $(\bar{Y}^w | D = 1) = 2.2$
- Average wage for control group: $(\bar{Y}^w | D = 0) = 1$
- Average causal effect: $(\bar{Y}^w | D = 1) - (\bar{Y}^w | D = 0) = 1.2$

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w	TE_i	D	OC_i	Y_i^w
1	BC	BC	1	1	0	0	BC	1
2	BC	BC	1	1	0	1	BC	1
3	BC	WC	0	2	2	0	BC	0
4	BC	WC	0	2	2	1	WC	2
5	BC	WC	0	2	2	0	BC	0
6	BC	WC	0	2	2	1	WC	2
7	BC	WC	0	2	2	0	BC	0
8	BC	WC	0	2	2	1	WC	2
9	WC	WC	4	4	0	0	WC	4
10	WC	WC	4	4	0	1	WC	4

Estimating Average Treatment Effects With Bad Controls 1/2

- Average wage for treatment group, **in occupation BC**:

$$(\bar{Y}^w | D = 1, OC = BC) = 1$$

- Average wage for control group, **in occupation BC**:

$$(\bar{Y}^w | D = 0, OC = BC) = 0.25$$

- Average causal effect, **in occupation BC**:

$$(\bar{Y}^w | D = 1, OC = BC) -$$

$$(\bar{Y}^w | D = 0, OC = BC) = 0.75$$

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w	TE_i	D	OC_i	\bar{Y}_i^w
1	BC	BC	1	1	0	0	BC	1
2	BC	BC	1	1	0	1	BC	1
3	BC	WC	0	2	2	0	BC	0
4	BC	WC	0	2	2	1	WC	2
5	BC	WC	0	2	2	0	BC	0
6	BC	WC	0	2	2	1	WC	2
7	BC	WC	0	2	2	0	BC	0
8	BC	WC	0	2	2	1	WC	2
9	WC	WC	4	4	0	0	WC	4
10	WC	WC	4	4	0	1	WC	4

Estimating Average Treatment Effects With Bad Controls 2/2

- Average wage for treatment group, **in occupation WC**:
 $(\bar{Y}^w | D = 1, OC = WC) = 2.5$
- Average wage for control group, **in occupation WC**:
 $(\bar{Y}^w | D = 0, OC = WC) = 4$
- Average causal effect, **in occupation WC**:
 $(\bar{Y}^w | D = 1, OC = WC) - (\bar{Y}^w | D = 0, OC = WC) = -1.5$

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w	TE_i	D	OC_i	\bar{Y}_i^w
1	BC	BC	1	1	0	0	BC	1
2	BC	BC	1	1	0	1	BC	1
3	BC	WC	0	2	2	0	BC	0
4	BC	WC	0	2	2	1	WC	2
5	BC	WC	0	2	2	0	BC	0
6	BC	WC	0	2	2	1	WC	2
7	BC	WC	0	2	2	0	BC	0
8	BC	WC	0	2	2	1	WC	2
9	WC	WC	4	4	0	0	WC	4
10	WC	WC	4	4	0	1	WC	4

Summing Up

- We have treatment that we know has some positive effect on most of the population ($Y_{1i} - Y_{0i} = 2$) and null for the rest ($Y_{1i} - Y_{0i} = 0$). In a setting with random assignment the simple difference in groups is the average causal effect for the entire population: $(\bar{Y}^w|D = 1) - (\bar{Y}^w|D = 0) = 1.2$
- But after we control for occupation, a variable that is also affected by the treatment, the simple difference for each sub-group is contaminated by selection bias: $(\bar{Y}^w|D = 0, OC = BC) = 0.75$ and $(\bar{Y}^w|D = 0, OC = WC) = -1.5$
- This happens because the treatment affects the composition groups BC and WC in terms of wages.

Composition Effects 1/3

- Without the treatment the potential outcomes for occupation and wages are such that most low wages are in occupation BC. So the average wages for groups BC and WC are:
- $(\bar{Y}_0^w | OC = BC) = 0.25$
- $(\bar{Y}_0^w | OC = WC) = 4$

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w
1	BC	BC	1	1
2	BC	BC	1	1
3	BC	WC	0	2
4	BC	WC	0	2
5	BC	WC	0	2
6	BC	WC	0	2
7	BC	WC	0	2
8	BC	WC	0	2
9	WC	WC	4	4

Composition Effects 2/3

- With the treatment the potential outcomes for occupation and wages are such that most who had were low wages before the treatment are now in occupation WC. So the average wages for groups BC and WC are:
 - $(\bar{Y}_1^w | OC = BC) = 1$
 - $(\bar{Y}_1^w | OC = WC) = 2.5$

		Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w
1	BC	BC		1	1
2	BC	BC		1	1
3	BC	WC		0	2
4	BC	WC		0	2
5	BC	WC		0	2
6	BC	WC		0	2
7	BC	WC		0	2
8	BC	WC		0	2
9	WC	WC		4	4

Composition Effects 3/3

- And when comparing average potential outcomes within occupational groups we get:

$$(\bar{Y}_1^w | OC = BC) - (\bar{Y}_0^w | OC = BC) = 0.75$$

$$(\bar{Y}_1^w | OC = WC) - (\bar{Y}_0^w | OC = WC) = -1.5$$

- Notice that to show this last part we didn't need the treatment assignment and effective outcomes. This is why skips all of it (but I hope it adds a bit more intuition!).

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w
1	BC	BC	1	1
2	BC	BC	1	1
3	BC	WC	0	2
4	BC	WC	0	2
5	BC	WC	0	2
6	BC	WC	0	2
7	BC	WC	0	2
8	BC	WC	0	2
9	WC	WC	4	4

Connecting This Example With MM's Example

- To map this toy data to the Table 6.1 in the book:
- Keep one of each type of response to treatments (keep rows 1, 3, 9).
- Assume a constant treatment effects of \$500 for all.
- And change the values of wages for Y_0 to \$1000, \$2000, and \$3000.

	Y_{i0}^{oc}	Y_{i1}^{oc}	Y_{i0}^w	Y_{i1}^w
1	BC	BC	1	1
2	BC	BC	1	1
3	BC	WC	0	2
4	BC	WC	0	2
5	BC	WC	0	2
6	BC	WC	0	2
7	BC	WC	0	2
8	BC	WC	0	2
9	WC	WC	4	4

But Why Did We Just Spent So Much Time on This?

- Bad Controls are a very important topic for several reasons:
 1. Even though it is fairly simple (after you understand it!) it is a fairly new development in econometrics, and it's great to see how we continue to learn new (simple and useful) things.

But Why Did We Just Spent So Much Time on This?

- Bad Controls are a very important topic for several reasons:
 1. Even though it is fairly simple (after you understand it!) it is a fairly new development in econometrics, and it's great to see how we continue to learn new (simple and useful) things.
 2. To improve our causal identification correcting flawed analysis (this is what we have been doing throughout this course). When analysts/researchers or commentators disregard or are indifferent to such flaws, they are supporting or spreading BS.

But Why Did We Just Spent So Much Time on This?

- Bad Controls are a very important topic for several reasons:
 1. Even though it is fairly simple (after you understand it!) it is a fairly new development in econometrics, and it's great to see how we continue to learn new (simple and useful) things.
 2. To improve our causal identification correcting flawed analysis (this is what we have been doing throughout this course). When analysts/researchers or commentators disregard or are indifferent to such flaws, they are supporting or spreading BS.
 3. Last but not least it will help us uncover the BS in an important policy debate...

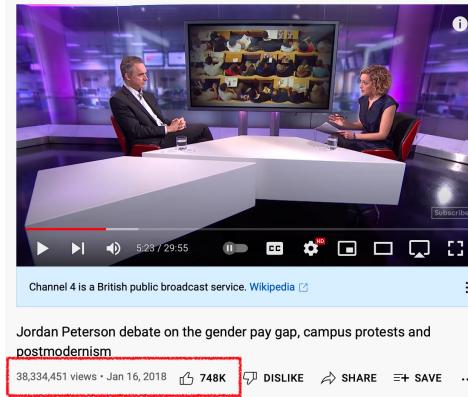
Example: Debate on the Gender Pay Gap

- Interviewer brings up the point that the gender pay gap in the UK is 9% as evidence that modern society is still primarily dominated by men. Her point is that gender **causes** a wage differential.
- Interviewee response:
 - "Multivariate analysis of the pay gap indicates that it doesn't exist"
 - "If you are a social scientist word your salt, you never do a univariate analysis [...]. You break it down by age, by occupation, interest, personality [and the gap disappears]"



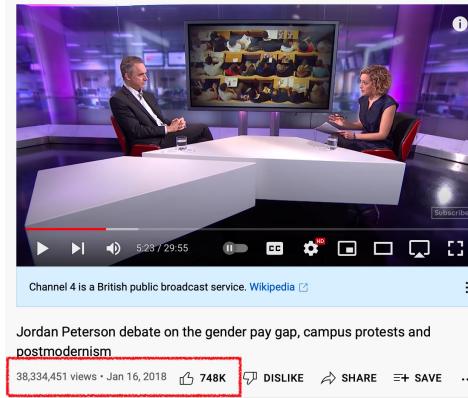
BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.



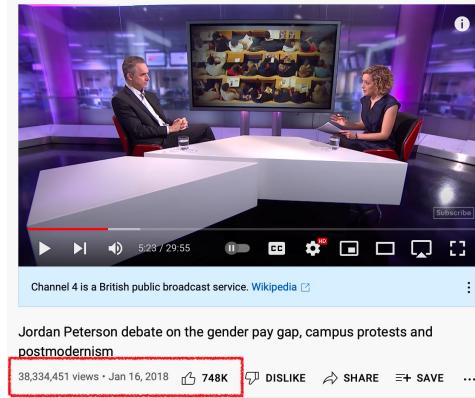
BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.



BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.
- Original statement



BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.
- Original statement
- Commentators critique:
 - multivariate
 - OVB



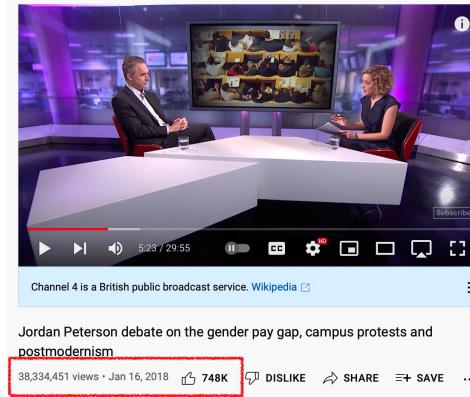
BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.
- Original statement
- Commentators critique:
 - multivariate
 - OVB
 - Can you point to where is the problem with?



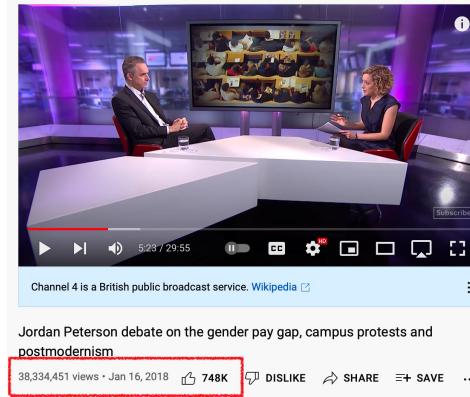
BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.
- Original statement
- Commentators critique:
 - multivariate
 - OVB
 - Can you point to where is the problem with?
- In addition to this point another likely point of BS has to do with the auxiliary and long regression in OVB for personality.



BS in The Argument That There is No Gender Wage Gap

- This course has given us the tools to clearly defined all the pieces a policy debates like this one.
- First, we identify all the elements, then we can clearly see where is the BS component.
- Original statement
- Commentators critique:
 - multivariate
 - OVB
 - Can you point to where is the problem with?
- In addition to this point another likely point of BS has to do with the auxiliary and long regression in OVB for personality.
- Why are we calling it BS and not wrong? Or a lie?



BS in The Argument That There is No Gender Wage Gap. Notes 1/2

- Original statement:
 - In English: the disproportionate domination of men in modern societies can be seen in the fact that gender causes a negative and substantial wage differential between men and women.
 - Using expectations and PO, and defining the treatment (D_i) as one if the individual's gender is female and zero if men:

$$\mathbb{E}(\ln Y_i | D_i = 1) - \mathbb{E}(\ln Y_i | D_i = 0) = -0.09 + \underbrace{\mathbb{E}(\ln Y_{i0} | D_i = 1) - \mathbb{E}(\ln Y_{i0} | D_i = 0)}_{\text{Selection bias}}$$

assuming **Selection Bias = 0**

- Using regression: $\ln Y_i = \alpha - 0.09D_i + e_i$. With the assumption that there is no OVB.

BS in The Argument That There is No Gender Wage Gap. Notes 2/2

- Commentator's critique:
 - In English: the cited gap does not compare individuals of similar age, occupation, personality and interests.
 - In regression terms: the cited gap comes from a regressor that fails to control for several variables leading to OVB.
 - For example, let's focus on the effect of excluding personality:

$$\text{long: } \ln Y_i = \alpha^l + \beta^l D_i + \gamma P_i + e_i^l$$

$$\text{aux: } P_i = \pi_0 + \pi_1 D_i + u_i$$

$$\text{OVB: } \beta^s - \beta^l = \pi_1 \gamma$$

- Following the commentator's rationale (from the domain of psychology), that women tend to be more agreeable than men: $\pi_1 > 0$
- And another rationale (from the domain of economics) that jobs with higher agreeableness tend to have lower wages: $\gamma < 0$
- Lead us to conclude that OVB is negative, which means that $\beta^s - \beta^l < 0$. Given that $\beta^s < 0$, this means that the true causal difference is lower (closer to zero).

BS in The Argument That There is No Gender Wage Gap. Notes 3/3

- In addition to this point another likely point of BS has to do with the auxiliary regression in OVB for personality.
 - This point is that his statement regarding $\pi_1 > 0$ are probably well grounded on research given his field of expertise. $\gamma < 0$ on the other hand requires expertise in personality and labor economics. And given that it is pretty hard to measure how personality traits affect earnings, we should be much more skeptical of the argument for γ (made with the same confidence as the first one, but probably reaching outside of field of expertise).
- Why are we calling it BS and not just wrong? Or a lie?
 - There is no reason to believe the commentator is lying. He does however, demonstrate a lack of interest towards the truth as explained above. He probably doesn't care about bad controls and how this might weakens his desired result (of no gender gap). This lack of interests about the truth, plus the fact that he is reaching out of his field of expertise are the two key characteristics that Frankfurt identify behind BS.

Acknowledgments

- MM