**Week 3 Lab Handout**

**PA 5032 – Applied Regression**

February 5th, 2021

Wenchen Wang

**PART A: F-Test Review ~10-15 min**

**PART B: Omitted and Irrelevant Variables~30-35 min**

**PART C: Report 1 Questions?**

*Download Lab3\_Data.dta from Canvas and load it into STATA.*

CPS, 46426 Observations, covering year 2020 from May to December.

**Variables.**

cpsidp

age

female

black

nchild

earnweek

hourwage

covidunaw

**Part A: Review of the F-Test**

Remember that an F Test is a test of joint significance, and it has different uses:

1. **Model Fit:** Does the model explain significantly more variation in the dependent variable than the mean of the dependent variable (a horizontal line)?
2. **Joint Significance**: Do a set of independent variables jointly significantly impact the dependent variable?
3. **Difference Between Two X Variables:** Are the impacts of two different independent variables significantly different from one another?

Suppose we regress workers unbale to work due to COVID-19 pandemic status on weekly earnings, controlling for the hourly wage.

*reg covidunaw earnweek hourwage*

A screenshot of a document

Description automatically generated

What is the interpretation of the F-Statistic and the related p-value?

We’re a little suspicious that the effect of weekly earnings on covid working status may be the same as the effect of hourly wage. What would be the null hypothesis of our F Test?

Test your null hypothesis with an F Test. What do the results show?

*test earnweek = hourwage*

*A white background with black text

Description automatically generated*

**Part B: Omitted and Irrelevant Variables**

We’re interested in looking at the relationship between covid not-working status and weekly earnings of individuals. Let’s start with a simple regression with covid not-working status as our dependent variable and weekly earnings as the independent variable.

*reg covidunaw earnweek*

A screenshot of a computer

Description automatically generated

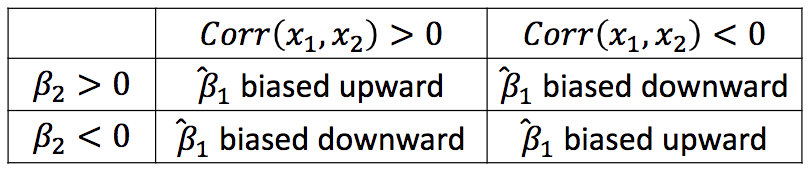
Describe the STATA output. What’s the relationship between covid not-working status and weekly earnings? What might be wrong with our model?

Even though there are a number of other things we could have controlled, think about the variable *black* (1=reference person identifies as black and 0=they identify as a different race). The fact that we did not control for this means that the estimates in our first model are likely biased.

In order to determine the direction of the bias, we have to determine the relationship between black and covid not-working status, as well as between black and weekly earnings. What is the direction of the bias in our first model, given that we did not control for black? Use the clues below to determine the bias direction.

A white background with black text

Description automatically generated



Now, run a regression controlling for black and check whether your prediction was correct.

*reg covidunaw earnweek*

A close-up of a number

Description automatically generated

Let’s shift our attention to what might affect workers not being able to work during covid. Start by regressing union on covid not-working status.

*reg union covidunaw*

*A screenshot of a computer code

Description automatically generated*

Your colleague suggests that this model is too simple and thinks you should add age and number of children as controls.

*reg union covidunaw age nchild*

*A screenshot of a computer

Description automatically generated*

Do you think that adding these controls has improved the model? Why or why not?

**Part C: Questions on Report 1?**