Econ 313, Fall 2021

HW #3

due October 7

For this assignment, you will apply panel data estimators to macroeconomic data. These estimators are very widely used in both microeconomics and macroeconomics.

Here are some class notes on panel data methods There are ample materials online discussing panel methods and almost any econometric text will cover them. No extra mathematical background is ordinarily required.

A *panel* is a dataset that displays both time-series and cross sectional variation. An example might be GDP for twenty different countries from 1980-2020. The # of cross sections is denoted N and the number of time periods is denoted T. If there are no missing observations, then the sample size will be TN.

Here is an example of panel data.

Table 1: Misconduct of NFL Teams

Team	Year	Murders	Caught Cheating
Patriots	2013	3	0
Patriots	2014	0	1
Everyone Else	2013	0	0
Everyone Else	2013	0	0

Note that the number of observations, 4, is too small to do any regression analysis, The data also suffer from measurement error as they exclude all of the Patriots misconduct for which they were not caught. There are 6000 unsolved murders in the U.S. each year. It is possible (likely?) that they were mostly committed by members of that criminal enterprise.

For this assignment, you may use the panel provided on covid-19 cases and stock returns. Students who instead construct their own panel, or who append the one provided, are eligible to receive up to 25% of their score as a bonus depending on the quality and originality of their dataset. Using a different, already available, dataset does not count.

0. If you are not using the dataset provided, describe your datset and why you chose it.

1. What is the "N" and what is the "T" for your dataset? Is it a balanced or unbalanced panel?

2. The command "xtset country month" tells Stata that your are working with a panel where "country" refers to the name of your cross sectional indicator and "month" is your time-series indicator. Run this command and then report some descriptive statistics. The ones reported with the "summarize" command are sufficient.

3. OLS with panel data is known as "pooled" OLS. Run pooled OLS in Stata using your panel. For the dataset that I provided, your goal is to estimate the impact of covid cases on stock returns. For now, exclude any lagged independent variables. Report your results and explain if any of the sources of misspecification covered in ECO 255 or ECO 313 are likely present.

4. Do you think that your model from #3 would benefit from including lagged independent variables? Explain. Experiment with different lag structures and provide the best possible specification.

With a panel, specific dates or countries may explain "unobserved heterogeneity." For states, California may behave systematically different than Kentucky in ways that are not explained by your independent variables. Another example is that 2008 may be systematically different than 2015 in ways that are not explained by your independent variables. If your panel displays unobserved heterogenity, then pooled OLS is misspecified.

5. Do you think that your model from #3 may exhibit unobserved heterogeneity? Explain.

6. If your unobserved heterogeneity is correlated with your independent variables, then the solution is to run a "fixed effects estimator." This is usually the case and simply entails adding dummy

variables for each country (or state, county, etc.) or date. The command "xtreg y x, fe r" will run a fixed effects estimator with dummy variables by cross section only. Run this specification for your model and report the results. Note that Stata does not automatically report the coefficients on the associated dummy variables.

7. If you want to add fixed effects for each time period, you have to do it manually in Stata. A command that will have fixed effects for time effects only is "reg y, x i.date, r." Note that you are not using "xtreg" here. Run this specification and report the results.

8. Combine #6 and #7 to write a command that includes fixed effects for both cross-section and time. Report the results.

9. Less often, your unobserved heterogeneity may be uncorrelated with your independent variables. Here, the solution is instead to run a "random effects estimator." A random effects estimator does not add dummy variables to your model but does re-compute your standard errors. The command "xtreg y x, re r" will run this estimator. Run this specification and report the results.

10. Among #4-#9, which do you think is the best econometric specification?

11. Suppose that I run pooled OLS when I should run fixed effects. Why will my regression coefficients be biased?

Bonus. A "Hausman test" formally tests between random and fixed effects. Describe this test and use it to chose between your fixed and random effects models.