Soc 73994, Homework #8
Panel Data & Multilevel Models

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised October 25, 2024

All answers should be submitted through Canvas. Be sure your response includes your name, the date, and a clear title, e.g. Homework # 8. If there is a huge amount of output for any analyses you run yourself, you may want to be selective in what you copy and paste into your assignment (but make sure you include enough so it is clear what commands you executed, e.g. you might show all the commands but only parts of the output).

This assignment focuses on basic panel data methods. The first two problems are required. The third problem is optional.

1. This example is adapted from the Stata 14 documentation on the `xtprobit` command. We have (synthetic) data on whether workers complain to managers at fast-food restaurants. The covariates are age (in years of the worker), grade (years of schooling completed by the worker), south (equal to 1 if the restaurant is located in the South, 0 otherwise), tenure (the number of years spent on the job by the worker), gender (of the worker; 1 = male, 0 = female), race (of the worker; 1 = Other, 2 = Black, 3 = White), income (in thousands of dollars by the restaurant), genderm (gender of the manager; 1 = male, 0 = female), chicken (1 = specializes in chicken, 0 = specializes in other types of food).

Note that we do not have multiple years of data for each restaurant. Instead, we have data for multiple employees for each restaurant. The term "cross-sectional time series," or xt, is a little misleading because the xt commands work fine in many cases when the data are not longitudinal. For example, you could have a sample of schools, with multiple students from each school.

Run the following code. You can add other commands if you wish.

```
webuse chicken, clear
label define sex 0 "Female" 1 "Male"
label values gender genderm sex
label define race 1 "Other" 2 "Black" 3 "White"
label values race race
keep complain age grade south tenure gender race income genderm chicken restaurant
xtset
xtsum
xtlogit complain age grade i.south tenure i.gender i.race income i.genderm i.chicken, nolog fe
est store fe
xtlogit complain age grade i.south tenure i.gender i.race income i.genderm i.chicken, nolog re
est store re
estimates table fe re, star
estimates restore re
margins south gender race genderm chicken
margins, dydx(south gender race genderm chicken)
```

Now answer the following questions.

a. Suppose that you were primarily concerned with omitted variable bias. What model might you favor, and why?
b. Suppose your primary concern was in assessing whether or not restaurants that specialize in chicken have more complaints than other types of restaurants. What model would you prefer then?
c. Even though they were specified on the command line, several variables are not included in the fixed effects model. Several cases are dropped too. Explain why. Use the results from the xtsum command to support your argument.
d. Interpret the results from the random effects model. What factors affect the likelihood of workers complaining? Use the results from both the `xtlogit` and `margins` commands. Run additional analyses if you think it would be helpful, e.g. you could run nested models using the KHB command.

2. Use the `clogit` and `melogit` commands to replicate and compare the results you got above for the fixed effects and random effects models. The `melogit` estimates will differ slightly from the `xtlogit` results. Specifically, run something like

```
clogit … [finish the command]
est store clogit
melogit … [finish the command]
est store melogit
estimates table fe clogit re melogit, star
```

3. (Optional but recommended if you are using Panel Data or multilevel data in your paper.) Do similar analyses using a data set of your choice. You don't have to perfectly mirror the above analysis but see if a random effects and/or fixed effects model can offer you any helpful insights.