

## PS207 Homework #5: Panel and TSCS Data

There is a Stata dataset on my website called `ps207hw5`. This is a dataset examining voter turnout in 49 US states (Louisiana is omitted because of an unusual election in 1982) plus the District of Columbia over 11 elections. In other words, this dataset is TSCS data on 50 units over 11 time periods. The variables in this dataset are:

*year*: The year of the election.

*stcode*: The ICPSR state code number.

*state*: The state name.

*vaprate*: The turnout rate as a percentage of the voting age population.

*midterm*: A dummy variable for midterm election years.

*regdead*: The number of days before the general election by which an individual needs to register.

*gsp*: State per capita income in 1000s of dollars.

There are also 9 regional dummy variables.

Download this dataset and do the following three problems. Turn in your writeups for these problems as well as the log files.

### Problem 1

Regress turnout as a percent of voting age population on the number of days before the general election by which an individual needs to register, state per capita income, the dummy variable for midterm elections, and the dummy variables for West North Central, the South, and the Border states. Describe the results of this regression.

### Problem 2

Obviously, we assumed that pooling our TSCS data was valid in Problem 1. Now let us test this assumption by estimating a fixed-effects model. We first must declare our data to be panel data. Type `xtset stcode year` to let Stata know that our units are indicated by *stcode*, and our time periods are indicated by *year*.

Then run a fixed-effects model by typing `xtreg vaprate midterm regdead gsp WNCentral South Border, fe`. The bottom line of the estimation results gives us an F test for pooling — can we conclude the 50 unit-specific dummy variables are all equal to zero? Is pooling appropriate in light of the results of this test? Why did some variables drop out of our estimation?

### Problem 3

Concerned with what happened to some variables in Problem 2, you decide to try a random-effects model. Type `xtreg vaprte midterm regdead gsp WNCentral South Border, re` to get a random effects model. Why do these results differ from those in Problem 2?

Type `xttest0`. What does this specification test tell you about the appropriateness of the random-effects model?

Finally, type `xthausman`. What does this test tell you? Discuss the tradeoffs between using pooled OLS, fixed-effects, and random-effects for this model.