

14.32: Econometrics

Problem Set 5

due Monday, November 17, 2025 at 1pm

1. True, False, Uncertain with Explanation:

- (a) Consider the following fixed effects regression model

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it},$$

where $\alpha_1, \dots, \alpha_n$ are entity fixed effects. Also consider the following regression model with “entity-demeaned” variables

$$\tilde{Y}_{it} = \beta_2 \tilde{X}_{it} + \tilde{u}_{it},$$

where $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$ and $\bar{Y}_i = (1/T) \sum_{t=1}^T Y_{it}$, and \tilde{X}_{it} and \tilde{u}_{it} are defined similarly. Assume $E[u_{it}] = 0$, but $E[u_{it} \mid X_{i1}, \dots, X_{iT}, \alpha_i] \neq 0$. The OLS estimator for $\hat{\beta}_1$ from the first regression is exactly the same as the OLS estimator for $\hat{\beta}_2$ from the second regression.

- (b) You estimate the following panel regression model:

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}.$$

Therefore, $\hat{\beta}_1$ captures the effect of X_{it} on Y_{it} after controlling for any shocks that are common to all entities in a given time period.

- (c) In the fixed effects regression model, you should always exclude one of the binary variables for the entities independent of whether an intercept is present in the equation or not.
- (d) Suppose you want to control for time-invariant factors like geography in a panel model of trade flows across countries. Including country fixed effects will allow you to identify the effect of time-invariant regressors such as distance between countries.
2. Policy debates and personalities get much attention in presidential elections. You will examine whether economic conditions also matter. More generally, what is the effect of economic conditions on the willingness to vote for an incumbent party? You will examine (2) using panel data on 50 states for the ten presidential elections from 1976 through 2012.

The incumbent party is the party of the sitting president. For example, in 2008, the incumbent party was Republican (Bush) and the candidates were McCain (Republican, lost) and Obama (Democrat, won). The panel data set also includes a binary variable indicating whether a given year was a non-recessionary year or not. Recessions are times of bad national economic conditions. Elections that overlapped with recession years were: 1976, 1984, 1992, and 2008. The remaining elections were in non-recession years. Variable *UR* stands for the unemployment rate in the state in October of the election year. *Income* is per capita income, in thousands of 2008 dollars.

- (a) Regressions (1) and (2) are identical, except that regression (1) uses heteroskedasticity-robust standard errors and regression (2) uses clustered standard errors, where the standard errors are clustered by state. Which is preferable? Explain.
- (b) Consider regression (4). What is the effect on the incumbent vote share of a one percentage point increase in the unemployment rate in a recession year, holding income constant? What is the effect on the incumbent vote share of a one percentage point increase in the unemployment rate in a non-recession year, holding income constant?
- (c) In regression (3), test the null hypothesis that the incumbent vote share does not depend on the unemployment rate (whether in a recession year or not), against the alternative that it does, at least in some years.
- (d) Consider regressions (4) and (5). What does regression (5) capture that regression (4) does not? Based on the results in the table, would you prefer specification (4) or (5)? Explain.
- (e) Provide an example of a variable not explicitly in regression (4) which could cause omitted variable bias, but which is controlled for by including time fixed effects.
- (f) Provide an example of a variable not explicitly in regression (4) which could cause omitted variable bias, but which is controlled for by including state fixed effects.
- (g) Explain, in everyday terms, what conclusions you draw from the regressions about the effect of economic conditions on the willingness of voters to support (or reject) the incumbent party.

Regression Results: Determinants of Voting for Incumbent Party, 1976-2012

Data are for U.S. states ($n = 50$) for $T = 10$ presidential election years (1976, 1980, \dots , 2012).

Outcome variable is the share of votes submitted for the incumbent party.

| | (1) | (2) | (3) | (4) | (5) |
|-------------------------------------------------------------------------------|-----------------|-----------------|------------------|------------------|------------------|
| UR | 0.13 (0.18) | 0.13 (0.25) | -0.73* (0.31) | -0.66* (0.30) | -2.83* (1.36) |
| $UR \times$ non-recession | | | 1.91** (0.64) | 1.98** (0.64) | 0.15+ (0.08) |
| Income | | | | 0.22* (0.09) | 0.20* (0.10) |
| UR^2 | | | | | 2.65 (1.89) |
| $UR^2 \times$ non-recession | | | | | -0.19 (0.12) |
| State fixed effects? | yes | yes | yes | yes | yes |
| year effects? | yes | yes | yes | yes | yes |
| Standard errors | HR | cluster | cluster | cluster | cluster |
| Number of observations | 500 | 500 | 500 | 500 | 500 |
| F-statistics testing that the coefficients on groups of variable are all zero | | | | | |
| $UR, UR \times Non - rec$ year | | | 4.58 (.015) | 4.83 (.012) | 5.35 (.001) |
| $UR^2, UR^2 \times$ Non-rec year | | | | | 1.97 (.151) |
| $UR, UR \times$ Non-rec year, $UR^2, UR^2 \times$ Non-rec year | | | | | 2.74 (.039) |
| Year dummy variables | 241.3 (.000) | 241.3 (.000) | 272.0 (.000) | 290.1 (.000) | 219.5 (.000) |

Notes: All regressions include an intercept. Standard errors are given in parentheses under estimated coefficients, and p -values are given in parentheses under F - statistics. Standard errors and F -statistics are heteroskedasticity-robust (HR) for regression (1) and are clustered for regressions (2)-(5), where clustering is at the state level. Coefficients are individually statistically significant at the +10%, *5%, ** 1% significance level.

3. We will think in this problem about effects of unemployment of the crime rate. Use data file `CRIME2.dta` for this exercise. It contains data on crime and unemployment for 46 cities for 1982 and 1987. There are 92 total observations

- (a) Run a regression of crime rate (crimes per 1000 people, variable *crmrate*) on unemployment (rate in %, variable *unem*) on a subsample restricted to the data for 1982 only, then repeat the same for the 1987 subsample. Do the results make sense? What is the sign and magnitude of coefficients would you expect to get?
- (b) Suggest an unobserved variable (or variable not in the data set) which varies across cities but plausibly varies little, or not at all, over time, and which plausibly could cause omitted variable bias in regressions (a).
- (c) Run a regression of crime rate on unemployment with city fixed effects on the full data set. Do the results change? If so, which set of results is more credible? Write explicitly what regression you are running.
- (d) Suggest a variable that may cause omitted variable bias, that can be fixed by adding time fixed effects to the last regression.
- (e) Do the results change when you add a time fixed effect to the regression that already has city fixed effects? If so, which set of regression results is more credible and why? Write explicitly what regression you are running.

Hint: the command `reghdfe` is useful in Stata for running regressions with multiple sets of fixed effects. You will need to install it by running `ssc install reghdfe`.

- (f) What conclusions would you draw about the effect of unemployment on crime?