

Problem Set 2: Panel Data and Safety Belt Legislation

This empirical portion of this problem set is based off of a data set used in Anderson (2008). That paper examined the effects of light trucks on traffic fatalities, but in doing so it also collected data on primary and secondary seat belt laws as additional covariates. This problem set will focus on the question of whether primary seat belt laws save lives. (A primary belt law stipulates that law enforcement can ticket a driver for not wearing a seat belt, regardless of whether he has broken any other laws. A secondary belt law stipulates that law enforcement can only ticket a driver for not wearing a seat belt if the driver has already been pulled over for simultaneously breaking a different traffic law.)

Unlike the data from the previous problem set, the data here have already been cleaned, with missing values dropped or replaced. The labels on the variables should be self-explanatory. We will first begin with a couple of theoretical exercises, however.

1. Question 10.3 from Wooldridge: For $T = 2$ consider the standard unobserved effects model:

$$y_{it} = \alpha + x_{it}\beta + c_i + u_{it} \tag{1}$$

Let $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ represent the fixed effects and first differences estimators respectively.

- (a) Show that $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. Hint: it may be easier to write $\hat{\beta}_{FE}$ as the “within estimator” rather than the fixed effects estimator.
- (b) Show that the standard errors of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. If you wish, you may assume that x_{it} is a scalar (i.e. there is only one regressor) and ignore any degree of freedom corrections. You are not clustering the standard errors in this problem.

2. Question 21-3 from Cameron-Trivedi (enhanced): Consider the fixed effects, two-way error component panel data model:

$$y_{it} = \alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \quad (2)$$

- (a) Show that the fixed effects estimator of β can be obtained by applying two within (one-way) transformations on this model. The first is the within transformation ignoring the time effects followed by the within transformation ignoring the individual effects. Assume the panel is balanced. (Hint: it may be easier to analyze the fixed effects regression using partialled regression.)
- (b) Show that the order of the transformations is unimportant. Give an intuitive explanation for why.
- (c) Does your answer to part (a) change if the panel becomes unbalanced (i.e., contains different numbers of observations for each

individual i). Why or why not?

3. We now begin with an actual analysis of the data. The goal here is to determine what effect, if any, primary belt laws have on the log of traffic fatalities per capita (we log the LHS variable because we believe the effect of safety belt laws should be proportional to the overall level of fatalities per capita).
 - (a) Run pooled bivariate OLS. Interpret. Add a quadratic time trend. Interpret. Add all covariates that you believe are appropriate. What happens when you add these covariates? Why?
 - (b) Ignore omitted variables bias issues for the moment. Do you think the standard errors from above are right? Compute the Huber-White heteroskedasticity robust standard errors (e.g., “, robust”). Do they change much? Compute the clustered standard errors that are robust to within-state correlation (e.g., “, cluster(state)”). Do this using both the “canned” command and manually using the formulas we learned in class. Do the standard errors change much? Are you surprised? Interpret.
 - (c) Compute the between estimator, both with and without covariates. Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? Do you believe those conditions are met? Are you concerned about the standard errors in this case?

- (d) Compute the RE estimator. Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? What are its advantages or disadvantages as compared to pooled OLS?
- (e) Do you think the standard errors from RE are right? Compute the clustered standard errors. Are they substantially different? If so, why? (i.e., what assumption(s) are being violated?)
- (f) Compute the FE estimator using only primary and a quadratic in year as the covariates. Compute the normal standard errors and the clustered standard errors. If they are different, why?
- (g) Add the same range of covariates to the FE estimator that you did to the OLS estimator. Are the FE estimates more or less stable than the OLS estimates? Why?