Problem Set - Panel Data

1. (Stock and Watson 2012, Ex 10.1) Consider the regression output in table 1 below.

   (a) New Jersey has a population of 8.1 million people. Suppose that New Jersey increased the tax on a case of beer by \$1 (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

   (b) The drinking age in New Jersey is 21. Suppose that New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

   (c) Suppose that real income per capita in New Jersey increases by 1% in the next year. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 90% confidence interval for your answer.

   (d) Should time effects be included in the regression? Why or why not?

   (e) A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. how would you test this hypothesis? (Be specific about the specification of the regression and the statistical test you would use.

2. In a random effects model, define the composite error $u_{it} = a_i + e_{it}$, where var $(a_i) = \sigma_a$, a_i is uncorrelated with e_{it}, and the e_{it} have constant variance σ_e^2 and are serially uncorrelated. Define $v_{it} = u_{it} - \lambda \overline{u}_i$. Derive the value of λ for which the v_{it} are serially uncorrelated. Also show that for this value of λ the v_{it} have mean zero, and constant variance.

3. Wooldridge exercise 10.2

4. Wooldridge exercise 10.4

5. (C) use http://fmwww.bc.edu/ec-p/data/wooldridge/jtrain (see data description below). You will be estimating the impact of a job training grant received by firm i in year t ($grant_{it}$) on hours of job training per employee ($hrsemp_{it}$) using the following model

$$hrsemp_{it} = \beta_0 + \delta_1 d88 + \delta_2 d89 + \beta_1 grant_{it} + \beta_2 grant_{i,t-1} + \beta_3 \log(employ_{it}) + a_i + u_{it}$$

   (a) Check whether you have missing values for the variables you need in your estimation. You should drop these observations before you estimate your model.

Dependent variable: traffic fatality rate (deaths per 10,000).

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	−0.66* (0.29)	−0.64+ (0.36)	−0.45 (0.30)	−0.69* (0.35)	−0.46 (0.31)	−0.93** (0.34)
Drinking age 18				0.028 (0.070)	−0.010 (0.083)		0.037 (0.102)
Drinking age 19				−0.018 (0.050)	−0.076 (0.068)		−0.065 (0.099)
Drinking age 20				0.032 (0.051)	−0.100+ (0.056)		−0.113 (0.125)
Drinking age						−0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				−0.063** (0.013)		−0.063** (0.013)	−0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes
***F*-Statistics and *p*-Values Testing Exclusion of Groups of Variables**							
Time effects = 0			4.22 (0.002)	10.12 (< 0.001)	3.48 (0.006)	10.28 (< 0.001)	37.49 (< 0.001)
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 (< 0.001)		31.96 (< 0.001)	25.20 (< 0.001)
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and *p*-values are given in parentheses under the *F*-statistics. The individual coefficient is statistically significant at the +10%, *5%, or **1% significance level.

Table 1: The effect of drunk driving laws on traffic deaths.

(b) Transform the variables in the regression to their demeaned values. Estimate the above equation using fixed effects by running an OLS regression on your transformed data. Check how many firms are used in the FE estimation using -codebook fcode if e(sample)-. How many total observations would be used if each firm had data on all variables (in particular, $hrsemp$) for all three years?

(c) Which regressors are statistically significant?

(d) Interpret the coefficient on grant?

(e) Is $grant_{i,t-1}$ insignificant? Is this surprising? Explain.

(f) Do larger firms provide their employees with more or less training, on average? How big are the differences? (For example, if a firm has 10% more employees, what is the change in average hours of training?)

(g) Estimate the equation in (a) using OLS and include firm specific dummies. Do you get the same results as in (b)–(c)?

Data description:

The file contains data on the scrap rate and worker training in Michigan manufacturing firms in the years 1987, 1988 and 1989. The number of observations equals 471 (157 firms * 3 years). Besides that, the file includes the following variables:

Variable description:

1. year 1987, 1988, or 1989
2. fcode firm code number
3. employ # employees at plant
4. sales annual sales, $
5. avgsal average employee salary
6. scrap scrap rate (per 100 items)
7. rework rework rate (per 100 items)
8. tothrs total hours training
9. union =1 if unionized
10. grant =1 if received grant
11. d89 =1 if year = 1989
12. d88 =1 if year = 1988
13. totrain total employees trained
14. hrsemp tothrs/totrain
15. lscrap log(scrap)
16. lemploy log(employ)
17. lsales log(sales)
18. lrework log(rework)
19. lhrsemp log(1 + hrsemp)
20. lscrap_1 lagged lscrap; missing 1987
21. grant_1 lagged grant; assumed 0 in 1987
22. clscrap lscrap - lscrap_1; year > 1987
23. cgrant grant - grant_1
24. clemploy lemploy - lemploy[t-1]
25. clsales lavgsal - lavgsal[t-1]
26. lavgsal log(avgsal)
27. clavgsal lavgsal - lavgsal[t-1]
28. cgrant_1 cgrant[t-1]
29. chrsemp hrsemp - hrsemp[t-1]
30. clhrsemp lhrsemp - lhrsemp[t-1]