

Economics 472

Problem Set #8

(1) Load the *panel* baseball data (panelbb) from the course website. This expanded baseball data set contains a set of fixed effects for each team. In particular, for each team, the data set contains wins, era, batting average and number of homeruns hit from the 1998-2004 seasons. To load the data into STATA, type

```
infile wins batavg era homeruns yanks zona atlanta balt boston cubs wsox  
reds indians rockies tigers marlins astros royals angels dodgers brewers  
twins mets oakland phillies pirates padres giants mariner cards drays  
rangers bluejays expos using a:panelbb.txt
```

If you have loaded the data into STATA correctly, there should be 210 observations on 34 variables. The variables listed after era denotes the team dummy variables. For example, $yanks_{it} = 1$ when the observation corresponds to the yankees.

(1a) First, run the following regression which ignores the panel structure of the data and does not include team fixed effects:

$$Wins_{it} = \beta_0 + \beta_1 Batavg_{it} + \beta_2 Homeruns_{it} + \beta_3 ERA_{it} + u_{it},$$

with i denoting team and t denoting year (or season). Interpret your results and include the regression output with your completed problem set.

(1b), Now, run a second regression which includes the team fixed effects:

$$Wins_{it} = \alpha FixedEffects_i + \beta_0 + \beta_1 Batavg_{it} + \beta_2 Homeruns_{it} + \beta_3 ERA_{it} + u_{it}.$$

In the above, the “Fixed Effects” portion of the regression will be the dummy variable coefficients on ALL of the 30 team dummies and α represents the coefficients on these dummies. To implement this regression using stata, you will have to list each team in your **regress** command. (I would just cut and paste from your infile statement to make sure that you get all the teams).

Look CAREFULLY at your regression output. Have any of the team dummies been dropped from the regression? Why has this happened?

(1c) Run the same regression as in (1b), but this time, do not include the intercept in the regression equation. To suppress the intercept in STATA, simply add “,nocons” to the end of the **regress** statement after you have listed all of your explanatory variables.

(1d) Using the results in (1c) and your initial regression, *test the hypothesis* that all of the team fixed effects are the same:

$$H_o : \alpha_{yanks} = \alpha_{zona} = \dots = \alpha_{expos}.$$

Think about what the restricted regression is in this context, noting that when you sum over all the team dummy variables, the result must be 1. What do you conclude based on the results of your test?

(2) Stock and Watson, exercise 8.1, selected parts:

(a) *Hint:* Note that the fatality rate is measured as deaths per 10,000 people. So, to obtain the impact on a population of 8.1 million people, you will need to multiply the parameter by 810. The confidence interval for this effect will also need to take this into account. See the discussion regarding the solution of exercise 5.4 for a related derivation.

(b) See hint provided in (a).

(c) See hint provided in (a).

(d) *Hint:* On the lower portion of the table are a set of F-statistics and p-values for a variety of hypothesis tests of interest.

(e) *Hint:* Think about possible omitted variables bias.

(f) *Hint:* Think about estimating a regression model that would allow the effect of unemployment to vary if the state is in the West. (You might try and create a dummy variable and add an interaction term to implement the test).