**Homework assignment 2**

Q1. I have given you a panel data on wages (Wage data) in which N=334, T=3 years (1984-1986).

For each ID, the data is sorted by year. You need to create an ID and a ‘year’ variable which you may do in Excel.

|  |  |  |
| --- | --- | --- |
| **Columns** | **Variable name** | **Description** |
| C1 | Edu | Education in years |
| C2 | Hr | Work hours per year |
| C3 | Wage | Dollar wage per hour |
| C4 | Famearn | Family earnings in dollars per year |
| C5 | Self | Dummy for self-employed |
| C6 | Sal | Dummy for salaried |
| C7 | Mar | Dummy for married |
| C8 | Numkid | Number of children |
| C9 | Age |  |
| C10 | unemp | Local unemployment percentage |

We need to do a regression to understand the determinants of “natural log (wages)” that is {ln(wage)}.

We need to understand the effect of the following variables: age, edu, numkid, hr, mar, sal, self, unemp on the dependent variable.

1. Find the best linear regression model. Check for multicollinearity and take appropriate actions. Interpret the estimates and write a report on your findings. Interpret model fit, t-values, meaning of coefficients, collinearity diagnostics, tests for heteroscedasticity etc.
2. Develop a model to test if there are nonlinear effects for age. What do you conclude?
3. Should you include family earnings in the explanatory variables? Why or why not?
4. Using the same regression model as in Q1, run fixed effects models and random effects models

i.e., FIXEDONE, FIXEDTWO, RANONE, RANTWO.

Create a table of coefficients side-by side with significant coefficients shown in bold (you may do this in Excel).

1. What is the effect of panel data models on the coefficients. What parameters have changed and by what percentage?
2. We are especially interested in the effect of education on wages. How much (%) has this coefficient changed across the different models? What is the correct estimate of the effect of education on wages?