Ec504: Computer Exercise 1: Estimation of the General Linear Model

The purpose of this exercise is to familiarise you with the basics of estimation and interpretation of the general linear model. The economic issue tested in this exercise is to try to establish a relationship between earnings and the levels of education, experience and job tenure, (human capital earnings function).

Read in the Stata data set, cex1.dta, from the Ec504 web site.

The data contain 6225 observations on British employees in 1998

age	age
grosspay	usual gross weekly pay (pounds/week)
workhrs	usual work hours
hourpay	hourly pay (pounds/hour)
agelftsc	age when left school
agelftft	age left full time education
xper	years of labour market experience
yearsed	years of education
tenure	years with current employer
male	1 if male, 0 if female
female	1 if female, 0 if male

1. As a means of familiarising yourself with the data, graph the wage variable against age and find the means, maxima and minima of each variable. Do the data make intuitive sense?

2. Regress hourly pay on years of experience. Interpret the parameter estimates. Are the variables statistically significant? What are the 95% confidence interval around the estimates? How well does your model fit the data?

3. Now add years of education, job tenure and the female dummy variable to your regression. Interpret the parameter estimates. How has the coefficient on experience changed? What can you infer about the correlation between experience and the other variables from this? What is the estimated impact of one year more experience with the same firm on pay? Are the variables statistically significant? What are the 95% confidence interval around the estimates? How well does your model fit the data?

4. Some human capital theorists believe that experience should entered non-linearly in a regression, (why?). Add the square of experience to your regressors. (you can do this with the command

```
gen xper2=xper^2
```

in Stata)

How do your estimates change? Has the fit of the model improved? At how many years of experience do wages appear to be maximised?

Graph the residuals from this regression against the fitted values. What do you see?

5. Now regress the log of hourly pay on the same variables. Interpret your estimates.

Graph the residuals from this regression against the fitted values from this regression. What do you see now?

6. Does the logarithmic or levels specification fit best?

(If you need the geometric mean, type the command

means *variable names*

)