## Ec317 Labour Economics Problem Set 13

1. Download the dataset "labor4.dta" from the course web site. It contains an extract from the January 1990 Current Population Survey for the U.S. with 8,162 observations on men and women employed in the private sector. The dataset contains a variety of demographic and employment related variables; we will only use some of them in this problem set.

Your mission in this problem is to look at the earnings differences between men and women and to draw some conclusions as to the source of these differences. You will have quite a bit of freedom to think about how you want to carry out your analysis. You should organize your completed assignment into three different parts: a "data description" section that gives an overview of the data and the differences between men and women; a second section that looks at different regression models and explores the effect of a dummy variable for women; and a third section that explores the issue whether the gender differential diverges over the career.

(a) Data Description Section

Prepare a table that gives means of hourly wages for men and women. Prepare a second descriptive table that looks at differences in other characteristics between men and women which might affect earnings (education, age, potential experience, weekly hours, full-time vs. part-time work, etc.). How big are these differences? Are the differences in personal characteristics likely to explain the differences in earnings between men and women? Look at the differences in educational attainment for those under 40 years of age, and those over 40. Is there any trend?

(b) Simple Regression Models

To estimate the (approximate) percentage difference in male and female wages, you can simply run a one variable regression of the log of wages on the dummy variable for females, i.e. in Stata

. reg lnwage female

Now include the standard human capital variables in your regression which also includes the dummy variable for females, i.e. in Stata

## . reg lnwage educ exp expsq female

(Notice that there is no variable **exp** or **expsq** on your dataset, you will have to construct them yourself.) Prepare a table of regression results where you explore other different specifications. Feel free to use the other variables provided on the dataset. Be creative and generate your own new variables (for example, interactions of preexisting variables). Write a few paragraphs explaining why you have chosen your particular specifications and why they are interesting. Explain why the size of the estimated gender effect differs between your different specifications.

(c) Models that Allow the Female Differential to Vary

In this section we want to explore the question whether the male/female difference in wages diverges or converges over the life-cycle. One way to do this is to include a series of dummy variables for females in different experience groups, e.g.

```
. gen female1 = female == 1 & exp <= 5
. gen female2 = female == 1 & exp >= 6 & exp <= 10
...</pre>
```

If you have created, for example, the variables female1 to female10 in this order, you can include them in your regression simply by typing

```
. reg lnwage educ exp expsq female1-female10
```

The coefficient on female1 now gives you the wage differential for women with less than 5 years of experience, the coefficient on female2 for those with 6 to 10 years of experience, etc. (in (b) above we were implicitly assuming that all of these are the same). Prepare again a table with alternative specifications. Write a few paragraphs explaining your results and interpreting their implications for the male/female wage gap. What could the results mean? Can you come up with an explanation based on human capital explaining your findings in (c)?