ECON 272 - 01, 02

Homework 5

Due Date: March 27

Note: Q.1 and Q.2 of this homework uses the same dataset as *Stata* lab 3: econ272_stata_lab3.

Q.1. We will test the hypothesis of whether married women earn more or less on the labour market.

a) Estimate the following fully saturated regression specification:
$Wage_i = \beta_0 + \beta_1 Married_i + \beta_2 Female_i + \beta_3 Female_i \times Married_i + \epsilon_i$

Use your *Stata* estimates to interpret β_0 and β_3. The outcome variable in the data is *incwage*.

b) An important omitted variable is the time worked by an individual. Define $WkWork_i$ as weeks worked in the year. Discuss how the estimate of β_1 would change if you include $WkWork_i$ in your estimation in a). Consider two cases: one where $Corr(Female_i, WkWork_i) > 0$ in the sample; second, where $Corr(Female_i, WkWork_i) < 0$, and discuss how the estimated β_1 would change in each case.

c) Now estimate this in the data by including weeks worked in your estimation (*wkswork1*). Would you say the estimate of β_1 in a) was biased upwards, or downwards? What does this mean about the sign of $Corr(Female_i, WkWork_i)$. Verify this in the data.

d) A second omitted variable is whether the worker has a college degree or not (*high_ed*). How does the β_2 and β_3 coefficients change upon including this variable? Use your regression coefficients and the omitted variable bias formula to infer whether women in the labour market are more likely to have a college degree.

d) Expand your set of covariates to include the worker's gender (*female*), age (*age*), and age-squared (*sq_age*). Also add in controls for worker location, geography, and industry. You can use the global macros for this. Report and interpret the coefficient estimates for each of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$, as defined in a). Would you say there a large or a small gap in earnings between married and unmarried male workers?

e) Use the *test* command in *Stata* to test the following null hypothesis: H_0: $\beta_1 + \beta_3 = 0$. Can you reject the null at the 1% level? What information is conveyed by $\beta_1 + \beta_3$?

f) Now re-estimate the same regression, but after taking the natural log of the outcome variable. Based on your results, is there a gap in earnings for married females, relative to married males?

g) Estimate the following regression:

$$\ln(Wage_i) = \beta_0 + \beta_1 Female_i + \beta_2 Age_i + \beta_3 Married_i + \beta_4 Age_i \times Married_i + \beta_5 Age_i \times Female_i + \delta \mathbf{X}_i + \epsilon_i$$

In g) include all the controls from d), as well as *College*, but exclude age-squared, and the interaction term $Married_i \times Female_i$. Interpret the β_2 and β_4 coefficients. Based on the coefficient estimates for β_4 and β_5, are there differential returns for female workers for an additional year of experience (assuming age and experience to be equivalent)?

h) Based on the coefficients in g), how much more would a female worker expect to earn from working 1 additional year? Using the *test* command in *Stata*, can you reject the null hypothesis: H_0: $\beta_2 + \beta_4 = 0$ with 95% confidence?

Q.2 We want to assess whether women in the labour force in metropolitan areas are more or less likely to finish college. Consider the fully saturated specification:

$$Pr(College_i = 1) = \beta_0 + \beta_1 Female + \beta_2 Metro_i + \beta_3 Female \times Metro + \epsilon_i$$

Estimate this in the data to assess whether women in the labour market in metropolitan areas have a higher or lower likelihood of finishing college. What information is provided by the coefficients β_0 and β_2?

Note: the relevant variables in the dataset are *metro*, *female* and *high_ed*.

Q.3. Consider the government thinking about a stimulus plan to boost the economy. A key part of the stimulus plan is to issue checks to household with a high propensity to consume. The government considers testing the hypothesis that families with more children have greated spending propensity. The population regression function is:

$$\ln(Consumption_i) = \beta_0 + \beta_1 \ln(Income_i) + \beta_2 \ln(Income_i) \times OneChild_i + \beta_3 \ln(Income_i) \times TwoChild_i + \beta_4 \ln(Income_i) \times ThreeChild_i + \beta_5 OneChild_i + \beta_6 TwoChild_i + \beta_7 ThreeChild_i + \delta \mathbf{X}_i + \epsilon_i$$

Income refers to the annual income of household i, while *Consumption* is the household's consumption. ln refers to the natural log function. *OneChild* is a binary variable equaling 1 if the family has only 1 child; *TwoChild* is a binary variable equaling 1 if the household has 2 children; *ThreeChild* is a binary variable equaling 1 if the household has 3 or more children. Remaining covariates are included in $\mathbf{X}$.

Assume the following regression coefficients:
$\hat{\beta}_1 = 0.11$; $se(\hat{\beta}_1) = .089$
$\hat{\beta}_2 = 0.31$; $se(\hat{\beta}_2) = .067$
$\hat{\beta}_3 = 0.61$; $se(\hat{\beta}_3) = .143$
$\hat{\beta}_4 = 0.21$; $se(\hat{\beta}_4) = .158$

a) Do changes in income have a large or small impact on consumption changes for families with no children?

b) Interpret the β_2 coefficient.

c) Based on the above evidence, as a policy maker, which type of families would you target when considering a stimulus package (recall, the goal is to increase household consumption).

Note: this problem is motivated from the expansion of the Earned Income Tax Credit during the 2007-09 recession. You can read more on this at: https://taxpolicycenter.org/briefing-book/what-earned-income-tax-credit