Econ 272

Homework 5

Name: Gianluca Crescenzo

Question 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:

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Married}_i + \beta_2 \text{Female}_i + \beta_3 (\text{Female}_i \cdot \text{Married}_i) + \epsilon_i.$$

Use your Stata estimates to interpret β_0 and β_3.

Solution. Running the regression gives:

```
------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
     married |   42988.52   1347.147    31.91   0.000      40348.1    45628.94
      female |  -12441.77   1507.421    -8.25   0.000    -15396.33   -9487.216
married_female |  -32428.07   1911.748   -16.96   0.000     -36175.1   -28681.03
       _cons |   76068.82   1091.899    69.67   0.000     73928.69    78208.94
------------------------------------------------------------------------------
```

$\widehat{\beta_0}$ is the average wage of single men. $\widehat{\beta_3}$ is how much the effect of marriage for women differs from the effect of marriage for men.

(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 the two cases: one where $\text{Corr}(\text{Married}_i, \text{WkWork}_i) > 0$ in the same, and second where $\text{Corr}(\text{Married}_i, \text{WkWork}_i) < 0$. Discuss how the estimated β_1 would change in each case.

Solution. If there is a positive correlation, omitting WkWork_i biases β_1 upward. Adding WkWork_i to the regression will reduce the estimated marriage effect. If there is a negative correlation, omitting WkWork_i biases β_1 downward. Adding WkWork_i will increase the estimated marriage effect.

(c) Now estimate this in the data by included 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 $\text{Corr}(\text{Married}_i, \text{WkWork}_i)$? Verify this in the data.

Solution. Running the regression gives:

```
------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
     married |   36375.78   1321.461    27.53   0.000      33785.7    38965.85
      female |  -12671.93   1471.949    -8.61   0.000    -15556.96   -9786.898
married_female |  -28534.23   1868.226   -15.27   0.000    -32195.97    -24872.5
```

```
     wkswork1 |   1904.327   36.28418    52.48   0.000     1833.21    1975.444
        _cons |  -10659.59   1966.592    -5.42   0.000   -14514.12   -6805.057
------------------------------------------------------------------------------
```

So part (a) was biased upwards. This also means $\text{Corr}(\text{Married}_i, \text{WkWork}_i) > 0$.

(d) A second omitted variable is whether the worker has a college degree or not. How do 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 labor market are more likely to have a college degree.

Solution.

(e) Expand your set of covariates to include the workers 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 $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\beta}_3$

Solution. Running the regressions gives:

```
--------------------------------------------------------------------------------
           wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------+---------------------------------------------------------------
        married |   30408.47   1330.407    22.86   0.000     27800.86    33016.07
         female |  -19937.97   1480.498   -13.47   0.000    -22839.76   -17036.19
 married_female |  -26237.66   1839.744   -14.26   0.000    -29843.57   -22631.75

                ...

          _cons |  -94117.24    7603.57   -12.38   0.000    -109020.3   -79214.19
--------------------------------------------------------------------------------
```

For β_1: holding all other variables constant, a married man earns about \$30,408 more per year than a comparable single man. For β_2: holding all else constant, a single woman earns about \$19,938 less per year than a comparable single man. For β_3: holding all else constant, the wage effect of marriage for women is \$26,238 lower than the wage effect of marriage for men.

(f) Use the *test* command to 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 $\beta_1 + \beta_3$?

Solution. Using the test command gave:

```
 ( 1)  married + married_female = 0

       F(  1, 56417) =    9.95
            Prob > F =    0.0016
```

Since the p-value is below 0.01, we can reject the null hypothesis. $\beta_1 + \beta_3$ gives the overall effect of marriage for women.

(g) 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?

Solution. Running the regression gives:

```
------------------------------------------------------------------------------
        ln_wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------+----------------------------------------------------------------
        married |   .2948742   .0101798    28.97   0.000     .2749216    .3148267
         female |  -.2099117   .0113078   -18.56   0.000    -.2320751   -.1877482
 married_female |  -.2258532   .0140448   -16.08   0.000    -.2533811   -.1983253

            ...

          _cons |   7.529791   .0593628   126.84   0.000     7.413439    7.646142
------------------------------------------------------------------------------
```

There is a large gap in earning for married females relative to married males.

(h) Estimate the following regression:

$$\ln \text{Wage}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + \beta_3 \text{Married}_i + \beta_4(\text{Age}_i \cdot \text{Married}_i) + \beta_5(\text{Age}_i \cdot \text{Female}_i) + \delta X_i + \epsilon_i.$$

In (g) include all the controls from (d), as well as College, but exclude age-squared, and the interaction term $\text{Married}_i \cdot \text{Female}_i$. Interpret the β_1 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)?

Solution. Running the regression gives:

```
------------------------------------------------------------------------------
        ln_wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------+----------------------------------------------------------------
         female |   -.189308   .0265723    -7.12   0.000      -.24139    -.137226
            age |   .0097006   .0005617    17.27   0.000     .0085997    .0108014
        married |   .2739934   .0267897    10.23   0.000     .2214854    .3265014
    age_married |  -.0032269   .0006104    -5.29   0.000    -.0044233   -.0020305
     age_female |  -.0035948   .0005943    -6.05   0.000    -.0047595     -.00243

                ...

          _cons |   8.146538   .0314906   258.70   0.000     8.084816    8.208259
------------------------------------------------------------------------------
```

β_1 shows that a female worker has a wage that is 18.9% lower than a comparable male. β_4 shows how the return to an additional year of age changes when a worker is married. Since $\beta_5 < 0$, there are differential returns, as each additional year of experience raises female log wages less than it does for similar males.

(i) 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?

Solution. Note that $\beta_2 + \beta_5 = 0.0061058$. For an unmarried female, each additional year of experience is associated with about a 0.61% increase in wages. The *test* command gives:

```
F(  1, 51267) =  182.97
         Prob > F =    0.0000
```

Thus we can reject the null hypothesis.

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

$$P(\text{College}_i = 1) = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Metro}_i + \beta_3 \text{Female}_i \cdot \text{Metro}_i + \epsilon_i.$$

Estimate this in the data to assess whether women in the labor 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?

Solution. Stata gives:

```
------------------------------------------------------------------------------
     high_ed | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      female |   .0678869   .0111446     6.09   0.000     .0460434    .0897305
       metro |   .0289121   .0028355    10.20   0.000     .0233546    .0344696
female_metro |  -.0076668   .0040858    -1.88   0.061     -.015675    .0003415
       _cons |   .2687816   .0077311    34.77   0.000     .2536286    .2839346
------------------------------------------------------------------------------
```

β_0 is the predicted probability that $\text{College}_i = 1$ for men who are not in a metropolitan area. β_2 is how the predicted probability that $\text{College}_i = 1$ changes if a man instead lives in a metropolitan area.

Question 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 households with a high propensity to consume. The government considers testing the hypothesis that families with more children have greater spending propensity. The population regression functions is:

$$\begin{aligned} \ln(\text{Consumption}_i) = \beta_0 + \beta_1 \ln(\text{Income}_i) &+ \beta_2 \ln(\text{Income}_i) \cdot \text{OneChild}_i \\ &+ \beta_3 \ln(\text{Income}_i) \cdot \text{TwoChild}_i \\ &+ \beta_4 \ln(\text{Income}_i) \cdot \text{ThreeChild}_i \\ &+ \beta_5 \text{OneChild}_i \\ &+ \beta_5 \text{TwoChild}_i \\ &+ \beta_5 \text{ThreeChild}_i + \delta X_i + \epsilon_i. \end{aligned}$$

Income_i refers to the annual income of household i, and Consumption_i is the households consumption. OneChild, TwoChild, and ThreeChild are binary variables. Assume the following regression coefficients:

$$\widehat{\beta}_1 = 0.11, \ \ \text{SE}(\widehat{\beta}_1) = .089$$
$$\widehat{\beta}_2 = 0.31, \ \ \text{SE}(\widehat{\beta}_2) = .067$$
$$\widehat{\beta}_3 = 0.61, \ \ \text{SE}(\widehat{\beta}_3) = .143$$
$$\widehat{\beta}_4 = 0.21, \ \ \text{SE}(\widehat{\beta}_4) = .158$$

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

Solution. For every 1% increase in income, there is a 0.11% increase in consumption. So income has a very small impact on consumption changes for families with no children.

(b) Interpret the β_2 coefficient.

Solution. Having exactly one child increases the consumption-income elasticity by 0.31 relative to households with no children.

(c) Based on the above evidence, as a policymaker, which type of families would you target when considering a stimulus package?

Solution. Since $\widehat{\beta}_3$, corresponding to families with two children, has the highest consumption-income elasticity, I would target those families when considering a stimulus package.