Homework no. 3

Econ 351, Introductory Econometrics

Instructor: Susumu Imai

Due date: Nov. 28, 2007.

100 points total

Question 1

The following model allows the return to education to depend upon the total amount of both parents’ education, called pareduc.

\[ \log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{educ} \times \text{pareduc} + \beta_3 \text{exper} + \beta_4 \text{tenure} + u. \]

i) Show that, in decimal form, the return to another year of education in this model is: \( \Delta \log(wage) / \Delta \text{educ} = \beta_1 + \beta_2 \text{pareduc} \). What sign do you expect for \( \beta_2 \)? Why? (10 points)

Take the derivative of \( \log(wage) \) with respect to education. The expected sign is positive, because higher parental education should increase the quality of education (quality of schools, etc).

ii) The estimation result is

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>5.65</td>
<td>0.13</td>
</tr>
<tr>
<td>education</td>
<td>0.047</td>
<td>0.010</td>
</tr>
<tr>
<td>educ*pareduc</td>
<td>0.00078</td>
<td>0.00021</td>
</tr>
<tr>
<td>experience</td>
<td>0.019</td>
<td>0.004</td>
</tr>
<tr>
<td>tenure</td>
<td>0.010</td>
<td>0.003</td>
</tr>
</tbody>
</table>

\( R^2 = 0.160, \; n = 722 \)

Interpret the coefficient on the interaction term. It might help to choose two specific values for pareduc, for example, pareduc=32 if both parents have a college education, or pareduc=24 if both parents have a highschool education, and to compute the estimated returns to educ.

Returns to education is higher for individuals whose parents have higher education. Returns to education for highschool graduated parents is 0.065, but the same for individuals with college educated parents are 0.072. (10 points)

iii) When pareduc is added as a separate variable to the equation, we get

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4.94</td>
<td>0.38</td>
</tr>
<tr>
<td>Education</td>
<td>0.097</td>
<td>0.027</td>
</tr>
<tr>
<td>pareduc</td>
<td>0.033</td>
<td>0.017</td>
</tr>
<tr>
<td>Educ*pareduc</td>
<td>0.0016</td>
<td>0.0012</td>
</tr>
<tr>
<td>Experience</td>
<td>0.020</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Does the estimated return to education now depend positively on parent education? Test the null hypothesis that the returns to education does not depend on parent education.

Since the interaction term is still positive, but not significantly different from zero. This is because the t-stat is 0.0016/0.0012=1.33 which is less than 1.980. Hence, the hypothesis that the returns to education does not depend on parent education cannot be rejected at 5% significance level. (10 points)

Question 2

Consider the following multiple regression model.

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \]

All the assumptions for classical linear models (which gives you unbiasedness, and the test statistics to be t and F). The sample size is \( n \). Suppose the researcher wants to test the hypothesis \( H_0 : \beta_1 = 3\beta_2 \) with 5% significance. Let \( c_5 \) be such that \( P(|t_{n-4}| > c_5) = 0.05 \) where \( t_{n-4} \) has a t-distribution with degrees of freedom \( n - 4 \).

1) (10 pts) How would you test the hypothesis after you have run the following OLS.

\[ y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \hat{u}_i \]

where \( \hat{u}_i \)'s are the residuals of the OLS. You also know the variance covariance matrix of the OLS coefficients. Explain the procedure in detail. (10 points)

Derive the variance of \( \hat{\beta}_1 - 3\hat{\beta}_2 \) as follows:

\[ \text{Var}[\hat{\beta}_1 - 3\hat{\beta}_2] = \text{Var}[\hat{\beta}_1] - 6\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) + 9\text{Var}[\hat{\beta}_2] \]

Then, derive the t-stat: \( t = \frac{\hat{\beta}_1 - 3\hat{\beta}_2}{\sqrt{\text{Var}[\hat{\beta}_1 - 3\hat{\beta}_2]}} \) and reject the hypothesis if \( |t| > c_5 \).

2) (10 pts) Suggest another way to test the hypothesis where you change variables so that you don’t need to know the variance covariance matrix of the OLS coefficients, just the standard error of the OLS estimates. Hint: transform the equations in this format:

\[ y_i = \beta_0 + [\beta_1 - 3\beta_2] x_{i1} + 3\beta_2 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \hat{u}_i \] Then, manipulate the equation in a similar way like in the lecture note so that you have a term involving \( \beta_2 \) and another involving \( \beta_3 \) (10 points)

\[ y_i = \beta_0 + [\beta_1 - 3\beta_2] x_{i1} + \beta_2 [3x_{i1} + x_{i2}] + \beta_3 x_{i3} + \hat{u}_i \] Run the regression where the dependent variable is \( y_i \) and the independent variables are \( x_{i1}, 3x_{i1} + x_{i2}, x_{i3} \). Then, test the hypothesis that the coefficient on \( x_{i1} \) is zero. (20 pts)
3) Suppose you want to test the hypothesis \( H_0 : \beta_1 = \beta_2 - \beta_3 \). Explain briefly how you would do it with and without using the variance covariance matrix of the OLS coefficients.

Hint: transform the equations in this format: \( y_i = \beta_0 + [\beta_1 - \beta_2 + \beta_3] x_{i1} + \beta_2 x_{i2} - \beta_3 x_{i3} + \beta_3 x_{i2} + \beta_3 x_{i3} + \hat{u}_i \). Then, manipulate the equation. (10 points)

\[
\text{Var} \left( \hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3 \right) = \text{Var} \left( \hat{\beta}_1 \right) + \text{Var} \left( \hat{\beta}_2 \right) + \text{Var} \left( \hat{\beta}_3 \right) - 2 \text{Cov} \left( \hat{\beta}_1, \hat{\beta}_2 \right) - 2 \text{Cov} \left( \hat{\beta}_1, \hat{\beta}_3 \right) + 2 \text{Cov} \left( \hat{\beta}_2, \hat{\beta}_3 \right)
\]

Then, derive the t-stat: 

\[
t = \frac{\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3}{\sqrt{\text{Var} \left( \hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3 \right)}}
\]

and reject the hypothesis if \( |t| > c \).

\[
y_i = \beta_0 + [\beta_1 - \beta_2 + \beta_3] x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \hat{u}_i.
\]

Run the regression where the dependent variable is \( y_i \) and the independent variables are \( x_{i1}, x_{i2} - x_{i3} + x_{i3} \). Then, test the hypothesis that the coefficient on \( x_{i1} \) is zero. (20 pts)

Question 3
We obtain the following estimated equation.

Dependent variable: sleep:total minutes per week spent sleeping at night
totwrk: total weekly minutes spent working
age: age in years
educ: years of education.
male: gender dummy: 1 if male and 0 otherwise.

<table>
<thead>
<tr>
<th>variables</th>
<th>Coefficient estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>3,840.83</td>
<td>235.11</td>
</tr>
<tr>
<td>totwrk</td>
<td>-0.163</td>
<td>0.018</td>
</tr>
<tr>
<td>educ</td>
<td>-11.71</td>
<td>5.86</td>
</tr>
<tr>
<td>age</td>
<td>-8.70</td>
<td>11.21</td>
</tr>
<tr>
<td>age squared</td>
<td>0.128</td>
<td>0.134</td>
</tr>
<tr>
<td>male</td>
<td>87.75</td>
<td>34.33</td>
</tr>
</tbody>
</table>

\( n = 706, \ R^2 = 0.123, \ adj. R^2 = 0.117 \)

1) All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence? (10 points)

\[
t = \frac{87.75}{34.33} = 2.556 > 1.658 \approx t_{0.05}(706) \text{ where the t critical value is for the one tailed test. Hence, the hypothesis that men do not sleep more than women is rejected. But the result that men sleep 88 minutes more per week, i.e. 12 minutes per day is not a very large difference, hence significant but not a strong evidence.} \]
2) Is there a statistically significant tradeoff between working and sleeping? What is the estimated tradeoff? (10 points)
   The coefficient estimate on total work minutes is significantly negative. Hence, one additional minute of work reduces sleep by 0.163 minutes.

3) What other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping? (10 points)
   You need to run the regression where the dependent variable is sleep minutes but independent variables are totwrk, educ, and male.