Name:

Student Number:

Midterm 2

ECO3151

Winter 2012

Instructions:

1. Print your name and student number at the top of this midterm
2. No programmable calculators
3. You can answer in pencil or pen
4. This midterm consists of 6 short answer questions
5. Total marks: 50

Question 1

a) Given the STATA output provided in Table 1, where *prices* is house price in thousands of dollars, *bdrms* is the number of bedrooms, *lotsize* is the size of the lot in square feet, and *sqrft* is the size of the house in square feet.

a) Write the underlying population model. (2 marks)

$$ln(price) = \beta_0 + \beta_1 ln(lotsize) + \beta_2 ln(sqrt) + \beta_3 bdrms + u \quad (1)$$

b) Interpret the *bdrms* coefficient estimate? (3 marks)

Adding one bedroom (holding other factors constant) will increase the house price by 3.7%

c) Is the *lsqrft* coefficient economically significant? Justify your answer. (3 marks)

If the square footage increase by 10%, then the price will go up by 7%. I would think that this is economically significant, i.e. economically important.

d) Does *lsqrft* have a statistically significant effect on *lprice*? Make sure to clearly state the null and alterative hypotheses and the underlying identifying assumptions. (4 marks)

$$H_0: \quad \beta_2 = 0$$
$$H_a: \quad \beta_2 > 0$$

Under MLR1-6, we know that

$$t_{stat} \sim t_{n-k-1} \quad (2)$$
$$t_{stat} \sim t_{84} \quad (3)$$
$$(4)$$

since n=88 and k=3.

$$t_{stat} = 7.54 > 1.671 \quad (5)$$

therefore I reject H_0 in favor of H_a at the 5% level of significance; square footage has a positive impact on the price of a home.

0.5cm
Question 2
Let the true population model be represented by the following equation

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where β_1 is the causal parameter of interest. Define $\tilde{\beta}_1$ as

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \bar{z})y_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}$$

where $z_i = ln(1 + x_i^2)$.Under what condition will $\tilde{\beta}_1$ consistently estimate β_1. Show your work. (5 marks)

$$\begin{aligned}
plim(\tilde{\beta}_1) &= plim(\frac{\sum_{i=1}^{n}(z_i - \bar{z})y_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}) &(6)\\
&= plim(\frac{\sum_{i=1}^{n}(z_i - \bar{z})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}) &(7)\\
&= plim(\frac{\sum_{i=1}^{n}(z_i - \bar{z})\beta_0 + \sum_{i=1}^{n}(z_i - \bar{z})\beta_1 x_i + \sum_{i=1}^{n}(z_i - \bar{z})u_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}) &(8)\\
&= plim(\frac{\beta_0 \sum_{i=1}^{n}(z_i - \bar{z}) + \beta_1 \sum_{i=1}^{n}(z_i - \bar{z})x_i + \sum_{i=1}^{n}(z_i - \bar{z})u_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}) &(9)\\
&= plim(\beta_1 + \frac{\sum_{i=1}^{n}(z_i - \bar{z})u_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}) &(10)\\
&= plim(\beta_1 + \frac{\sum_{i=1}^{n}(z_i - \bar{z})(u_i - \bar{u})/n}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})/n}) &(11)\\
&= \beta_1 + \frac{E(z_i - E(z_i))(u_i - E(u_i))}{E(z_i - E(z_i))(x_i - E(x_i))} &(12)
\end{aligned}$$

See class answer for the necessary conditions

Question 3
Assume the following population model

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u \tag{13}$$

where *wage* is the hourly wage, *educ* years of education, and *exper* years of work experience.

a) Education and experience are probably correlated. Is that problematic for carrying out a hypothesis test of whether education affects the wage? Explain your answer (3 marks)

As long as there is no perfect collinearity one can still carry out a t-test

b) Given the STATA output in Table 2, test whether experience matters (statistically speaking) in the model. Make sure to clearly state the null and alterative hypotheses and the underlying identifying assumptions. (5 marks)

$$\begin{aligned} H_0: &\quad \beta_2 = \beta_3 = 0 \\ H_a: &\quad H_0 does not hold \end{aligned}$$

Under MLR1-6, we know that

$$F_{stat} \sim F_{q,n-k-1} \tag{14}$$
$$F_{stat} \sim F_{2,522} \tag{15}$$
$$\tag{16}$$

since n=526 and k=3.

$$F_{stat} = ((0.3003 - .1858)/2)/(1 - 0.3003)/522) = 42.7 > 3.07 \tag{17}$$

therefore I reject H_0 in favor of H_a at the 5% level of significance; experience matters.

Question 4 It was argued that having a large sample can be beneficial. explain. (5 marks)

A large sample will mean that you will have small standard errors. This is useful to carry out test. It also means that if your estimator is consistent, the large sample will mean that the odds of getting a guess that is close to the true are good.

Question 5 Let the true population model be represented by the following equation

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{18}$$

where β_1 is the causal parameter of interest. For each individual i (in the sample of size n) you can observe x_i, but not y_i. You can, however, observe y_i^* where

$$y_i^* = y_i + \epsilon_i \tag{19}$$

and where ϵ is a noise term that is unobserved.

Under what conditions would regressing y^* on x provide a consistent estimator of β_1? (5marks)

You would still the original MLR1-4 (where MLR-4 is $E(x|u) = 0$). You would also need that the noise not be correlated with the x.

Question 6

Assume the following model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{20}$$

where β_1, β_2 and β_3 are the causal parameters of interest. Evaluate the following claims:

i) *"The fact than $\hat{\beta}_2^{ols}$ has an upward bias, i.e. $E(\hat{\beta}_2^{ols}) > \beta_2$ means that the OLS coefficient estimate will always be greater than β_2."* (5 marks)

Not true. the actual guess could be above. below or equal to the true one. All you know is that on average your guesses will be high.

ii) *"The presence of multicolinearity between x_2 and x_3 (but not perfectly correlated) means that one cannot consistently estimate β_2."* (5 marks)

No, collinearity does not invalidate the MLR1-4 to hold.

iii) *"If $\hat{\beta}_2^{ols}$ is consistent (for β_2, it will generate an accurate guess of β_2."* (5 marks)

no, consistency means that the odds of being close to the true population parameter get better and better as n gets large and larger. So the odds of being close are good if you have a large sample.

Equation Sheet

- In the simple linear regression model where x is the explanatory variable, and y the dependent variable,

$$\begin{aligned}
\hat{\beta}_0^{ols} &= \bar{y} - \hat{\beta}_1^{ols}\bar{x} \\
\hat{\beta}_1^{ols} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
se(\hat{\beta}_1^{ols}) &= \frac{\hat{\sigma}}{\sqrt{SST_x}}
\end{aligned}$$

- In the multiple linear regression model

$$se(\hat{\beta}_j^{ols}) = \frac{\hat{\sigma}}{(SST_j(1 - R_j^2))^{1/2}}$$

- $R^2 = SSE/SST$
- $SST_y = \sum_{i=1}^n (y_i - \bar{y})^2$, $SSR_y = \sum_{i=1}^n (y_i - \hat{y})^2$, $SSE_y = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
- $\hat{\sigma}^2 = (\sum_{i=1}^n \hat{u}_i^2)/(n - k - 1)$
- $Var(z) = E(z - E(z))(z - E(z)) = E(z^2) - E(z)^2$
- $Cov(z, y) = E(z - E(z))(y - E(y)) = E(zy) - E(z)E(y)$
-

$$F = \frac{(R_{ur}^2 - R_u^2)/q}{(1 - R_{ur}^2)/n - k - 1}$$