Practice Exam 2
Econ471
Fall 2020
Solution

Part I. Multiple Choice Questions

1. The normality assumption implies that:
a. the population error *u* is dependent on the explanatory variables and is normally distributed with mean equal to one and variance σ^2.
b. the population error *u* is independent of the explanatory variables and is normally distributed with mean equal to one and variance σ.
c. the population error *u* is dependent on the explanatory variables and is normally distributed with mean zero and variance σ.
d. the population error *u* is independent of the explanatory variables and is normally distributed with mean zero and variance σ^2.
Answer: d

2. Consider the equation, $Y = \beta_1 + \beta_2 X_2 + u$. A null hypothesis, H_0: $\beta_2 = 0$ states that:
a. X_2 has no effect on the expected value of β_2.
b. X_2 has no effect on the expected value of Y.
c. β_2 has no effect on the expected value of Y.
d. Y has no effect on the expected value of X_2.
Answer: b

3. Which of the following statements is true?
a. If the calculated value of F statistic is higher than the critical value, we reject the alternative hypothesis in favor of the null hypothesis.
b. The F statistic is always nonnegative as SSR_r is never smaller than SSR_{ur}.
c. Degrees of freedom of a restricted model is always less than the degrees of freedom of an unrestricted model.
d. The F statistic is more flexible than the t statistic to test a hypothesis with a single restriction.
Answer: b

4. Which of the following statements is true?
a. The standard error of a regression, $\hat{\sigma}$, is not an unbiased estimator for σ, the standard deviation of the error, u, in a multiple regression model.
b. All estimators that are unbiased are also consistent.
c. Almost all economists agree that unbiasedness is a minimal requirement for an estimator in regression analysis.
d. All estimators in a regression model that are consistent are also unbiased.
Answer: a

5. In a multiple regression model, the OLS estimator is consistent if:
a. there is no correlation between the dependent variables and the error term.
b. there is a perfect correlation between the dependent variables and the error term.
c. the sample size is less than the number of parameters in the model.
d. there is no correlation between the independent variables and the error term.
Answer: d

6. If $\delta_1 = Cov(x_1,x_2) / Var(x_1)$ where x_1 and x_2 are two independent variables in a regression equation, which of the following statements is true?
a. If x_2 has a positive partial effect on the dependent variable, and $\delta_1 > 0$, then the inconsistency in the simple regression slope estimator associated with x_1 is negative.
b. If x_2 has a positive partial effect on the dependent variable, and $\delta_1 > 0$, then the inconsistency in the simple regression slope estimator associated with x_1 is positive.
c. If x_1 has a positive partial effect on the dependent variable, and $\delta_1 > 0$, then the inconsistency in the simple regression slope estimator associated with x_1 is negative.
d. If x_1 has a positive partial effect on the dependent variable, and $\delta_1 > 0$, then the inconsistency in the simple regression slope estimator associated with x_1 is positive.
Answer: b

7. Which of the following statements is true under the Gauss-Markov assumptions?
a. Among a certain class of estimators, OLS estimators are best linear unbiased, but are asymptotically inefficient.
b. Among a certain class of estimators, OLS estimators are biased but asymptotically efficient.
c. Among a certain class of estimators, OLS estimators are best linear unbiased and asymptotically efficient.
d. The LM test is independent of the Gauss-Markov assumptions.
Answer: c

8. Changing the unit of measurement of any independent variable, where log of the independent variable appears in the regression:
a. affects only the intercept coefficient.
b. affects only the slope coefficient.
c. affects both the slope and intercept coefficients.
d. affects neither the slope nor the intercept coefficient.
Answer: a

9. If a regression equation has only one explanatory variable, say x_1, its standardized coefficient must lie in the range:
a. -2 to 0.
b. -1 to 1.
c. 0 to 1.
d. 0 to 2.
Answer: b

10. Which of the following correctly represents the equation for adjusted R^2?
a. $\bar{R}^2 = 1 - [SSR/(n-1)]/[SST/(n+1)]$
b. $\bar{R}^2 = 1 - [SSR/(n-k-1)]/[SST/(n+1)]$
c. $\bar{R}^2 = 1 - [SSR/(n-k-1)]/[SST/(n-1)]$
d. $\bar{R}^2 = 1 - [SSR]/[SST/(n-1)]$
Answer: c

11. The following simple model is used to determine the annual savings of an individual on the basis of his annual income and education.
Savings = $\beta_0 + \partial_0$ Edu + β_1Inc+u
The variable 'Edu' takes a value of 1 if the person is educated and the variable 'Inc' measures the income of the individual. The inclusion of another binary variable in this model that takes a value of 1 if a person is uneducated, will give rise to the problem of _____.

a. omitted variable bias
b. self-selection
c. dummy variable trap
d. heteroskedastcity
Answer: c

Part 2. Quantitative Questions

1. The F-statistic for testing a set of linear hypotheses is given by the formula

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

where SSR_r is the sum of squared residuals from the restricted regression, SSR_{ur} is the sum of squared residuals from the unrestricted regression, q is the number of restrictions under the null hypothesis, and k is the number of regressors in the unrestricted regression. Prove that this formula is the same as the following formula based on the regression R^2 of the restricted and unrestricted regression:

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

Solution: Since $R^2 = 1 - \frac{SSR}{SST}$, SSR=SST(1- R^2). Substitution into the first equation then results in the second equation, once the "1" in the numerator is canceled, and the *SST* is factored out in the numerator.

2. Consider the following multiple regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

You want to consider certain hypotheses involving more than one parameter. You decide to test the joint hypotheses using the F-statistics. For each of the cases below specify a restricted model and indicate how you would compute the F-statistic to test for the validity of the restrictions.

(a) $\beta_1 = -\beta_2;\ \beta_3 = 0$

(b) $\beta_1 + \beta_2 + \beta_3 = 1$

(c) $\beta_1 = \beta_2/\beta_3$

Solution: (a) The restricted model is $Y_i = \beta_0 + \beta_1(X_{1i} - X_{2i}) + u_i$ and the *F*-statistic would be

$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(n-3-1)}$$

.

(b) The restricted model is $Y_i - X_{3i} = \beta_0 + \beta_1(X_{1i} - X_{3i}) + \beta_2(X_{2i} - X_{3i}) + u_i$ and the *F*-statistic would be

$$F = \frac{(SSR_r - SSR_{ur})/1}{SSR_{ur}/(n-3-1)}$$

(c) This is not a linear restriction. Hence you cannot use the *F*-test to test for its validity.

3.

(i) With $df = n - 2 = 86$, we obtain the 5% critical value from Table G.2 with $df = 90$. Because each test is two-tailed, the critical value is 1.987. The t statistic for H_0: $\beta_0 = 0$ is about -.89, which is much less than 1.987 in absolute value. Therefore, we fail to reject $\beta_0 = 0$. The t statistic for H_0: $\beta_1 = 1$ is $(.976 - 1)/.049 \approx -.49$, which is even less significant. (Remember, we reject H_0 in favor of H_1 in this case only if $|t| > 1.987$.)

(ii) We use the SSR form of the F statistic. We are testing $q = 2$ restrictions and the df in the unrestricted model is 86. We are given $SSR_r = 209{,}448.99$ and $SSR_{ur} = 165{,}644.51$. Therefore,

$$F = \frac{(209{,}448.99 - 165{,}644.51)}{165{,}644.51} \cdot \left(\frac{86}{2}\right) \approx 11.37,$$

which is a strong rejection of H_0: from Table G.3c, the 1% critical value with 2 and 90 df is 4.85.

(iii) We use the R-squared form of the F statistic. We are testing $q = 3$ restrictions and there are $88 - 5 = 83$ df in the unrestricted model. The F statistic is $[(.829 - .820)/(1 - .829)](83/3) \approx 1.46$. The 10% critical value (again using 90 denominator df in Table G.3a) is 2.15, so we fail to reject H_0 at even the 10% level. In fact, the p-value is about .23.

(iv) If heteroskedasticity were present, Assumption MLR.5 would be violated, and the F statistic would not have an F distribution under the null hypothesis. Therefore, comparing the F statistic against the usual critical values, or obtaining the p-value from the F distribution, would not be especially meaningful.

4. Write $y = \beta_0 + \beta_1 x + u$, and take the expected value: $E(y) = \beta_0 + \beta_1 E(x) + E(u)$, or $\mu_y = \beta_0 + \beta_1 \mu_x$ since $E(u) = 0$, where $\mu_y = E(y)$ and $\mu_x = E(x)$. We can rewrite this as $\beta_0 = \mu_y - \beta_1 \mu_x$. Now, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. Taking the plim of this we have $\text{plim}(\hat{\beta}_0) = \text{plim}(\bar{y} - \hat{\beta}_1 \bar{x}) = \text{plim}(\bar{y}) - \text{plim}(\hat{\beta}_1) \cdot \text{plim}(\bar{x}) = \mu_y - \beta_1 \mu_x$, where we use the fact that $\text{plim}(\bar{y}) = \mu_y$ and $\text{plim}(\bar{x}) = \mu_x$ by the law of large numbers, and $\text{plim}(\hat{\beta}_1) = \beta_1$.

5. (i) If $\Delta cigs = 10$ then $\Delta \log\widehat{(bwght)} = -.0044(10) = -.044$, which means about a 4.4% lower birth weight.

(ii) A white child is estimated to weigh about 5.5% more, other factors in the first equation fixed. Further, $t_{white} \approx 4.23$, which is well above any commonly used critical value. Thus, the difference between white and nonwhite babies is also statistically significant.

(iii) If the mother has one more year of education, the child's birth weight is estimated to be .3% lower. This is not a huge effect, and the *t* statistic is only one, so it is not statistically significant.

(iv) The two regressions use different sets of observations. The second regression uses fewer observations because *motheduc* or *fatheduc* are missing for some observations. We would have to reestimate the first equation (and obtain the *R*-squared) using the same observations used to estimate the second equation.