Practice Exam 2
Econ471
Fall 2020

Part I. Multiple Choice Questions

d.

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.

2. Consider the equation, $Y = \beta_1 + \beta_2X_2 + u$. A null hypothesis, H_0: $\beta_2 = 0$ states that:

b.

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.

3. Which of the following statements is true?

b.

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.

a.

b.

4. Which of the following statements is true?
a. The standard error of a regression, $\hat{\sigma}^2$, is ~~not an~~ unbiased estimator for σ^2, the standard deviation of the error, u, in a multiple regression model.
b. All estimators that are unbiased are also consistent. may not converge.
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.

d.

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.

b.

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?

$\beta_1 + \delta_1 \beta_2$

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.

c

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.

a

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.

b.

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.

C

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. $\bar{R}^2 = 1 - [SSR]/[SST/(n-1)]$

d 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 虚假变量

d. heteroskedastcity 异方差

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 $SSRr$ 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:

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

$$R^2 = 1 - \frac{SSR}{SST}$$

$$SSR = (1 - R^2)SST.$$

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

$Y_i = \beta_0 + X_{1i} + \beta_2(X_{2i} - X_{1i}) + u_i$

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

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

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

$F = \dfrac{\_\_\_\_\_\_ /1}{\_\_\_\_\_\_ /(n-3-1)}$

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

(C) is not a linear restriction

can't use F-test.

3. In section 4.5 in the textbook, we used the following log-log model for testing the rationality of assessment of housing prices:

$$\log(price) = \beta_0 + \beta_1 \log(assess) + \beta_2 \log(lotsize) + \beta_3 \log(sqrft) + \beta_4 bdrms + u,$$

where

price = house price.
assess = the assessed housing value (before the house was sold).
lotsize = size of the lot, in square feet.
sqrft = square footage.
bdrms = number of bedrooms.

If the assessed housing price is a rational valuation, then a 1% change in *assess* should be associated with a 1% change in *price*.

We now use a level-level formulation to test for the rationality of the assessed valuation.

(i) In the simple regression model

$$price = \beta_0 + \beta_1 assess + u,$$

the assessment is rational if $\beta_1 = 1$ and $\beta_0 = 0$. The estimated equation is

$$\widehat{price} = -14.47 + 0.976\, assess$$

$$(16.27)\ (0.049)$$

$$n = 88,\ SSR = 165{,}644.51,\ R^2 = 0.820.$$

First, test the hypothesis that $H_0: \beta_0 = 0$ against the two-sided alternative. Then, test $H_0: \beta_1 = 1$ against the two-sided alternative. What do you conclude?

$H_0: \beta_0 = 0$: $t = \dfrac{-14.47}{16.27} = -0.889367 \Rightarrow$ accept.

df = 86. $|t| < 1.987$

$H_0: \beta_1 = 1$: $t = \dfrac{0.976 - 1}{0.049} = -0.489796 \Rightarrow$ accept.

$|t| < 1.987$

(ii) To test the joint hypothesis that $\beta_0 = 0$ and $\beta_1 = 1$, we need the SSR in the restricted model. This amounts to computing $\sum_{i=1}^{n}(price_i - assess_i)^2$, since the residuals in the restricted model are just $price_i - assess_i$.(No estimation is needed for the restricted model because both parameters are specified under H_0.) This turns out to yield SSR = 209,448.99. Carry out the *F* test for the joint hypothesis.

$$F = \frac{(209448.99 - 165644.51)/2}{165644.51/86} \approx 11.37$$

$\Rightarrow$ reject

(iii) Now, test $H_0: \beta_2 = 0,\ \beta_3 = 0,\ \beta_4 = 0$ in the model

$$\text{price} = \beta_0 + \beta_1 \text{assess} + \beta_2 lotsize + \beta_3 \text{sqrft} + \beta_4 bdrms + u.$$

The R-squared from estimating this model using the same 88 houses is 0.829.

$$F = \frac{(0.829 - 0.820)/3}{(1 - 0.829)/83} = 1.46$$

$\Rightarrow$ accept

(iv) If the variance of *price* changes with *assess*, *lotsize, sqrft*, or *bdrms*, what can you say about the *F* test from part (iii)?

heteroskedasticty not have F distribution under null hypothesis.

Can't use F test.

4. In the simple regression model $y = \beta_0 + \beta_1 x + u$ it was shown in class that under MLR.1 through MLR.4 assumptions the slope coefficient $\hat{\beta}_1$is consistent for β_1. Using $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, show that $plim\hat{\beta}_0 = \beta_0$. [You need to use the consistency of $\hat{\beta}_1$ and the law of large numbers, along with the fact that $\beta_0 = E(y) - \beta_1 E(x)$].

$$
\begin{aligned}
plim\,\hat{\beta}_0 &= plim\,(\bar{y} - \hat{\beta}_1 \bar{x}) \\
&= E(y) - E(x) \cdot plim\,\hat{\beta}_1 \\
&= E(y) - \beta_1 E(x). \\
&= \beta_0
\end{aligned}
$$

5. The following equations were estimated using the data in BWGHT.TXT:

$$\log(\widehat{bwght}) = 4.66 - .0044 cigs + .0093 \log(faminc) + .016 parity$$

b% = 100 x 0.0044 cigs.

$$(.22) \quad (.0009) \quad (.0059) \quad (.006)$$

$$+.027 male + .055 white$$

$$(.010) \quad (.013)$$

$$n = 1{,}388, \quad R^2 = 0.0472$$

and

$$\log(\widehat{bwght}) = 4.65 - .0052 cigs + .0110 \log(faminc) + .017 parity$$

$$(.38) \quad (.0010) \quad (.0085) \quad (.006)$$

$$+.034 male + .045 white - .0030 motheduc + .0032 fatheduc$$

$$(.011) \quad (.015) \quad (.0030) \quad (.0026$$

$$n = 1{,}190, \quad R^2 = 0.0493,$$

where

bwght = birth weight, in pounds.
cigs = average number of cigarettes the mother smoked per day during pregnancy.
parity = the birth order of this child.
faminc = annual family income.
motheduc= years of schooling for the mother.
fatheduc = years of schooling for the father.
male = dummay variable, equals 1 if child is male.
white = dummy variable, equals 1 if child is classified as white.

10 cig per days means 4.4% lower.

(i) In the first equation, interpret the coefficients on the variable *cigs*. In particular, what is the effect on birth weight from smoking 10 cigarettes per day?
(ii) How much more is a white child predicted to weigh than a nonwhite child, holding the other factors in the first equation fixed? Is the difference statistically significant?
(iii) Comment on the estimated effect and statistical significance of *motheduc*.
(iv) From the given equation, why are you unable to compute the *F* statistic for joint significance of *motheduc* and *fatheduc*? What would you have to do to compute the *F* statistic?

5.5% $\quad t = \dfrac{0.055}{0.013} = 4.23 > 1.77.$ => reject.
=> significant.

not significant.