STOCKHOLM UNIVERSITY
Department of Statistics
Mahmood Ul Hassan

EXAM IN ECONOMETRICS I
2022-05-30

Time: 5 hours

- Aids allowed: Pocket calculator, Textbook: Wooldridge, J.M. Introductory Econometrics: A Modern Approach, Cengage and notes written in the textbook are allowed.
- Note that no formula sheet is provided.
- Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.
- The exam consists of four questions. To score maximum points on a question, solution need to be clear, detailed, and well-motivated.

Question. 1 (5+4+3+3+5+5+=25 Points)

A study was made by a retail merchant to determine the relationship between weekly advertising expenditure and sale. The following data were recorded. X=Advertising Cost ($) Y=Sale ($)

X	Y	$\hat{u}$
40	385	-87.5381
20	400	-8.1218
25	395	-29.2259
20	365	-43.1218
30	475	34.6701
50	440	-64.7462
40	490	17.4619
20	420	11.8782
50	560	55.2538
40	525	52.4619
25	480	55.7741
50	510	5.2538

$$\sum X^2 = 15650, \quad \sum Y^2 = 2512925, \quad \sum XY = 191325$$

a) Estimate the model $Y = \beta_0 + \beta_1 X + u$ using the above information and interpret the estimated regression coefficient.

b) Assess how much variability in sale is not because of advertising cost.

c) Estimate the standard error of the regression ($\hat{\sigma}$).

d) Estimate the sale when advertising cost is 45$.

e) Construct 95% C.I for β_0.

f) Test the hypothesis that the weekly sale of a randomly selected week is 400$ when advertising cost is 40$.

Question. 2 (12+13=25 Points)

1) Given the model: $Y_i = \beta_1 X_i + \beta_2 X_i^2 + \mu_i$, where $\sigma_i^2 = \sigma^2 X_i^\delta$

   a) How do you transform the model to obtain homoscedastic errors for δ=2? Also show that the disturbance of the transformed model is homoscedastic.

   b) For δ=2, estimate the beta parameters efficiently using the data below

Y: 1 3 7 12 18

X: 1 2 4 5 8

2) From a regression analysis we collect the following information

Model: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \mu_t$

	Period-1	Period-2	Period 1+2
SST	46000	42000	108000
SSR	15000	12000	35000
N	20	20	40

The estimation results for the period 1 and period 2 are based on two independent regressions. The results from period 1+2 are obtained using all 40 observations.

Are the parameters of the model the same for the two periods? Perform a formal test. Your solution should include null- and alternative hypothesis in terms of the beta parameters, test statistic, its distribution with specification of degrees of freedom, decision rule, results, and a conclusion.

Question. 3 (7+12+6=25 Points)

a) The demand and supply equations are given below

$$\text{Demand Curve}: Q = \alpha_0 + \alpha_P P + u$$
$$\text{Supply Curve}: \; Q = \beta_0 + \beta_p P + v$$

find the reduced form for P and Q.

b) Consider the regression model $Y_i = \beta_1 + \beta_2 X_i + u_i$, where X takes the value 0 for group A and the value 1 for group B.

1) Show that the OLS estimates are $\hat{\beta}_1 = \bar{Y}_A$ and $\hat{\beta}_2 = \bar{Y}_B - \bar{Y}_A$.

2) Show that $\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 = \dfrac{n_A \, . \, n_B}{n_A + n_B}$

c) Let grad be a dummy variable for whether a student-athlete at a large university graduates in five years. Let hsGP A and SAT be high school grade point average and SAT score, respectively. Let study be the number of hours spent per week in an organized study hall. Using data on 420 student-athletes, we obtain:

$$\ln\left(\frac{p(grade=1)}{1-p(grade=1)} \right) = -1.17 + 0.24 hsGPA + 0.00058 SAT + 0.073 study$$

I. Holding hsGPA fixed at 5.0 and SAT fixed at 1000, compute the estimated difference in the graduation probability for someone who spent 14 hours per week in study hall and someone who spent 8 hours.

II. Compute the estimated graduation probability for someone having hsGPA 8.0 SAT score 1200 and who spent 20 hours per week in study hall.

Question. 4 (2.5+…+2.5=25 Points)

Indicate which alternative that is correct. Answering more than one alternative result in 0 points on the sub-question. No motivation is required.

1) Suppose we have estimated the regression model,

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + u_i$$

Let $\hat{y}_i$ be the fitted value of y_i for each i. Now, we estimate the artificial model,

$$\hat{y}_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + \gamma_1 \hat{y}_i + \gamma_2 \hat{y}_i^2 + v_i$$

to test $H_0 : \gamma_1 = \gamma_2 = 0$ against H_1: H_0 is wrong. Choose the correct statement.

a) H_1 can be equivalently rewritten as $H_1 : \gamma_1 \neq 0$ and $\gamma_2 \neq 0$.

b) An F-test cannot be appropriate for testing H_0.

c) This test is called the Augmented Dicky-Fuller test.

d) Rejection of H_0 suggests that there can be omitted variables.

e) None of the above is correct.

2) Consider the regression model,

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + u_i$$

where errors may be heteroskedastic. Choose the most incorrect statement.

a) The OLS estimators are consistent and unbiased.

b) We should report the OLS estimates with the robust standard errors.

c) The Gauss-Markov theorem may not apply.

d) The GLS cannot be used because we do not know the error variances in practice.

e) We should take care of heteroskedasticity only if homoskedasticity is rejected.

3) In order to estimate the wage equation, an econometrician regresses the log of wage on individual's observed characteristics including years of schooling. In this problem, however, the error term contains unobserved characteristics such as motivation, and it is likely that the error is correlated with the years of schooling, i.e., a highly motivated person tends to study more and also make more money. Choose the wrong statement.

a) The years of schooling is endogenous.

b) The OLS estimator is still consistent and asymptotically normal.

c) The OLS estimator is not unbiased.

d) The IV estimator can be used to estimate the coefficients.

e) None of the above is wrong.

4) In the model $y_i = \beta_1 + \beta_2 x_i + u_i$, with $i = 1, 2, \ldots, n,$ the error term is heteroscedastic when

a) $\text{var}(u_i) = 2\sigma^2 + 5$ for all $i = 1, 2, \ldots, n.$

b) $u_i = 20 + \varepsilon_i$ with $\text{var}(\varepsilon_i) = 10$ for all $i = 1, 2, \ldots, n.$

c) $\text{var}(u_i) = \frac{1}{5}\sigma_i^2$ for all $i = 1, 2, \ldots, n.$

d) $\text{var}(u_i) = \frac{\sigma^2}{10} + 5$ for all $i = 1, 2, \ldots, n.$

5) If two or more explanatory variables are highly correlated then which of the statement is not correct

a) Ordinary least squares estimate of the parameters can be estimated.

b) The ordinary least squares estimates are still unbiased.

c) The standard error of estimates becomes small.

d) The confidence interval becomes wider.

e) The probability of type II error increases.

6) Which of the following statements is TRUE concerning the standard regression model?

a) y has a probability distribution

b) x has a probability distribution

c) The disturbance term is assumed to be correlated with x

d) For an adequate model, the residual ($\hat{\mu}$) will be zero for all sample data points

7) Which of the following statement is true about the Quadratic model $y = \beta_o + \beta_1 x_1 + \beta_2 x_1^2 + u$

a) The model is linear in variable.

b) The model is linear in variable but non-linear in parameters.

c) The model is linear in parameter but non-linear in variable.

d) The model is non-linear in parameters and variable.

8) If we have a data set for which the dependent variable can only take on positive values and we observe many observations with $y_i = 0$. Which model should be used to obtain appropriate parameter estimates?

a) The Tobit model
b) OLS
c) Logistic regression
d) The Heckman error correction model

9) Simultaneous equation models are typically estimated with:

a) Regression models

b) 2SLS/IV regression

c) Logistic regression

d) Weighted least squares

10) Which the following is not related to multicollinearity problems:

a) Some of our independent variables are highly correlated.

b) We have a high variance inflation factor.

c) We have included a regressor that is uncorrelated with our dependent variable.

d) We get a high R^2 and high variance in our parameter estimates for the slopes.