STOCKHOLM UNIVERSITY
Department of Statistics
Mahmood Ul Hassan

EXAM IN ECONOMETRICS I
2022-04-28

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	$\hat{Y}$	$\hat{u}$
40	385	472.538	-87.5381
20	400	408.122	-8.1218
25	395	424.226	-29.2259
20	365	408.122	-43.1218
30	475	440.33	34.6701
50	440	504.746	-64.7462
40	490	472.538	17.4619
20	420	408.122	11.8782
50	560	504.746	55.2538
40	525	472.538	52.4619
25	480	424.226	55.7741
50	510	504.746	5.2538

$$\sum X = 410, \ \sum X^2 = 15650, \ \sum Y = 5445, \sum Y^2 = 2512925, \ \sum XY = 191325$$

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 because of advertising cost and interpret it.

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

d) For what value of advertising cost the expected value of sale is 485$.

e) Construct 95% C.I for β_1.

f) Test the hypothesis that the mean value of weekly sale is 500$ when advertising cost is 42$.

Question. 2 (12+13=25 Points)

1) Determine which of the following models are linear in parameter or linear in variables or both.
   1) Geometric or power curve model: $y_i = \beta_o x_i^{\beta_1} + u_i$
   2) Quadratic model: $y = \beta_o + \beta_1 x_1 + \beta_2 x_1^2 + u$
   3) $\ln(y_i) = \beta_o + \beta_1 \ln(x_{1i}) + u_i$
   4) $\ln(y_i) = \beta_o + \beta_1 \frac{1}{x_{1i}} + u_i$

2) The following equations were estimated using the data. The first equation is for houses with fireplaces and the second for houses without fireplaces

$$\hat{Y} = 64.9970 - 39.4931X_2 + 7.6578X_3 - 0.7553X_8$$

where n = 6, R^2 = 0.9839 and SSR = 2.74194.

$$\hat{Y} = 15.5566 + 14.4781X_2 + 0.9598X_3 - 0.1062X_8$$

where n = 18, R^2 = 0.6757 and SSR = 195.2586.

The third equation combine houses with and without fireplaces.

$$\hat{Y} = 17.37162 + 13.44330X_2 + 0.90053X_3 - 0.11457X_8$$

where n = 24, R^2 = 0.7146 and SSR = 236.5995.

Compute the usual Chow test for testing the null hypothesis that the regression equations are the same for houses with a fireplace and for houses without a fireplace.

Question. 3 (7+10+8=25 Points)

a) The two-equation system in "supply and demand form", that is, with the same variable y_t appearing on the left-hand side:

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$
$$y_1 = \alpha_2 y_2 + \beta_2 z_2 + u_2$$

If $\alpha_1 \neq 0, \alpha_2 \neq 0$ *and* $\alpha_1 \neq \alpha_2$, find the reduced form for Y_1.

b) Consider the simple regression model

$$y = \beta_o + \beta_1 x + u$$

and let z be a binary instrumental variable for x. Show that the IV estimator $\hat{\beta}_1 = \frac{\text{cov}(z,y)}{\text{cov}(z,x)}$ can be written as $\hat{\beta}_1 = \frac{(\bar{y}_1 - \bar{y}_o)}{(\bar{x}_1 - \bar{x}_o)}$

where $\bar{y}_o$ and $\bar{x}_o$ are the sample averages of y_i and x_i over the part of the sample with $z_i = 0$, and where $\bar{y}_1$ and $\bar{x}_1$ are the sample averages of y_i and x_i over the part of the sample with $z_i = 1$.

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:

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

I. Holding hsGPA fixed at 3.0 and SAT fixed at 1200, compute the estimated difference in the graduation probability for someone who spent 11 hours per week in study hall and someone who spent 5 hours.

II. Compute the estimated non-graduation probability for someone having hsGPA 5.0 SAT score 1500 and who spent 15 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) 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) $\text{var}(u_i) = \sigma_i^2$ for all $i = 1, 2, \ldots, n.$

   d) $\text{var}(u_i) = \frac{\sigma^2}{2} + 4$ for all $i = 1, 2, \ldots, n.$

2) Under the standard hypotheses of the General Linear Model $y_i = \beta_o + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i$, with $i = 1, 2, \ldots, n,$ efficiency in the Gauss-Markov sense of the OLS estimator of $\beta = (\beta_o, \beta_1 \ldots, \beta_k)$ means that:

   a) The variance of the OLS estimator of β is equal to one.
   b) There is no alternative linear and unbiased estimator of β with a smaller variance.
   c) The expected value of the OLS estimator of β is always equal to zero.
   d) The maximum likelihood estimator of β has smaller variance than OLS estimator of β.

3) Simultaneous equation models are typically estimated with:
   a) Regression models
   b) Logistic regression
   c) 2SLS/IV regression
   d) Weighted least squares

4) For the regression specification y = α + βx + ε the OLS estimates result from minimizing the sum of
   a) $(\alpha + \beta x)^2$ b) $(\alpha + \beta x + \varepsilon)^2$
   c) $(y - \alpha + \beta x)^2$ d) none of these

5) Suppose you are using the specification
   $wage = \alpha + \beta Education + \delta Male + \theta Education * Male + \varepsilon$
   In this specification the influence of Education on wage is the same for both males and females if

   (a) $\delta = 0$ (b) $\theta = 0$ (c) $\theta = \delta$ (d) $\theta + \delta = 0$

6) Correlation between the error term and an explanatory variable can arise because
   a) of error in measuring the dependent variable
   b) of a constant non-zero expected error
   c) the equation we are estimating is part of a system of simultaneous equations
   d) of multicollinearity

7) In the specification wage = β Education + δ Male + θ Female + ε
   a) there is perfect multicollinearity
   b) the computer will refuse to run this regression
   c) both a) and b) above
   d) none of the above

8) In general, omitting a relevant explanatory variable creates
   a) bias and increases variance
   b) bias and decreases variance
   c) no bias and increases variance
   d) no bias and decreases variance

9) Multicollinearity causes
   a) low R-squares
   b) biased coefficient estimates
   c) biased coefficient variance estimates
   d) none of these

10) Suppose your dependent variable is aggregate household demand for electricity for various cities. To correct for heteroskedasticity you should
   a) multiply observations by the city size
   b) divide observations by the city size
   c) multiply observations by the square root of the city size
   d) divide observations by the square root of the city size
   e) none of these