Name: _____

Free practice exam · examtex.com/free-exams

Introductory Econometrics — Practice Exam

Instructions: Part A: answer ALL 20 multiple-choice questions (2 points each). Each question has four options, A to D, of which exactly one is correct. Part B: answer ALL 4 questions (15 points each) and show your working. log denotes the natural logarithm. Standard normal critical values: 1.645 (5 percent one-sided or 10 percent two-sided), 1.96 (5 percent two-sided) and 2.576 (1 percent two-sided). A calculator may be used. Time allowed: 120 minutes.

Time Limit: 120 minutes

Multiple Choice (20 questions)

[2 points] 1. Using data on 120 cafés, a researcher estimates the OLS regression $\widehat{\text{revenue}} = 310 + 14.5\,\text{seats}$, where revenue is average daily revenue in euros and seats is the number of seats. A café with 40 seats has an average daily revenue of 820 euros. Its fitted value and residual are:

A. fitted value 890 euros; residual 70 euros

B. fitted value 890 euros; residual −70 euros

C. fitted value 580 euros; residual 240 euros

D. fitted value 820 euros; residual 0, because OLS residuals always sum to zero

[2 points] 2. Using monthly data, a city transport office estimates $\widehat{\log(\text{trips})} = 7.40 - 0.65\log(\text{fare}) + 0.30\log(\text{income})$, where trips is the number of taxi trips, fare is the average fare and income is average household income. Holding income constant, a 2 percent rise in the average fare is associated with:

A. a fall of about 1.3 percent in the number of trips

B. a fall of about 0.65 percent in the number of trips

C. a fall of about 130 percent in the number of trips

D. a fall of about 0.013 percent in the number of trips

[2 points] 3. Using data on 500 primary schools, a researcher estimates $\widehat{\text{score}} = 410 + 24\log(\text{income})$, where score is the school's average test score in points and income is the average household income of its catchment area in euros. A school whose catchment income is 10 percent higher is predicted to have an average score that is higher by approximately:

A. 24 points

B. 0.24 points

C. 2.4 points

D. 2.4 percent

[2 points] 4. In a regression of log(price) for used cars on age, mileage and a dummy variable automatic (1 if the car has an automatic gearbox, 0 if manual), the coefficient on automatic is 0.25. Holding age and mileage constant, by exactly what percentage is the predicted price of an automatic car HIGHER than that of a manual car?

A. 25.0 percent
B. 28.4 percent
C. 22.1 percent
D. 2.5 percent

[2 points] 5. A researcher has quarterly data on a shop's sales and creates four dummy variables, Q1, Q2, Q3 and Q4, one for each quarter of the year. She regresses sales on advertising spending, an intercept and ALL FOUR quarter dummies. What happens?

A. OLS runs normally, but the four dummy coefficients are biased towards zero
B. The errors become heteroskedastic, so robust standard errors are required
C. The R^2 is forced to equal 1
D. There is perfect multicollinearity, so OLS has no unique solution

[2 points] 6. For a sample of delivery drivers, OLS gives $\widehat{\text{pay}} = 9.00 + 1.20\,\text{exper} + 2.50\,\text{city} + 0.40\,(\text{city} \times \text{exper})$, where pay is hourly pay in euros, exper is years of experience and city = 1 for drivers working in a large city and 0 otherwise. The estimated effect of one more year of experience on the hourly pay of a CITY driver is:

A. 1.60 euros
B. 1.20 euros
C. 4.10 euros
D. 0.40 euros

[2 points] 7. A regression of annual earnings, measured in euros, on years of schooling gives a slope of 2,400 with a standard error of 300 and an R^2 of 0.18. Earnings are then re-measured in thousands of euros and the regression is run again. The new results are:

A. slope 2.4, standard error 300, t-statistic 0.008, R^2 0.18
B. slope 2.4, standard error 0.3, t-statistic 8, R^2 0.00018
C. slope 2.4, standard error 0.3, t-statistic 8, R^2 0.18
D. slope 2,400,000, standard error 300,000, t-statistic 8, R^2 0.18

[2 points] 8. A researcher adds one more regressor to a multiple regression. In the larger model, the new variable's coefficient has a t-statistic of 0.60. Compared with the smaller model, the larger model's R^2 and adjusted R^2 ($\bar{R}^2$):

A. both rise, because the new variable adds explanatory power
B. R^2 falls and $\bar{R}^2$ rises
C. both fall, because the new variable is not statistically significant
D. R^2 rises and $\bar{R}^2$ falls

[2 points] 9. In a regression with 800 observations, the coefficient on a training-programme dummy is 0.054 with a standard error of 0.030. Theory says that training cannot lower productivity. At the 5 percent significance level, the null hypothesis $H_0 : \beta = 0$ is:

A. rejected against both the one-sided alternative $\beta > 0$ and the two-sided alternative $\beta \neq 0$

B. rejected against the one-sided alternative $\beta > 0$, but not against $\beta \neq 0$

C. not rejected against either alternative, because t is below 1.96

D. rejected against $\beta \neq 0$, but not against the one-sided alternative $\beta > 0$

[2 points] 10. Using a large sample, the 95 percent confidence interval for a slope coefficient β_1 is $[-0.10,\ 0.70]$. Which statement is correct?

A. $H_0 : \beta_1 = 0$ is rejected at the 5 percent level (two-sided)

B. There is a 95 percent probability that the true β_1 lies between -0.10 and 0.70

C. The OLS estimate is 0.30 and its standard error is 0.40

D. $H_0 : \beta_1 = 0.75$ is rejected at the 5 percent level (two-sided)

[2 points] 11. A regression with an intercept and 4 regressors is estimated on $n = 105$ observations and has $R^2 = 0.20$. The F-statistic for the null hypothesis that all four slope coefficients are zero is:

A. 6.25

B. 25.0

C. 0.25

D. 6.50

[2 points] 12. An agronomist regresses wheat yield per hectare on fertilizer use per hectare across farms, but has no measure of soil quality. Better soil raises yields, and farmers with poorer soil apply MORE fertilizer to make up for it. The OLS estimate of the effect of fertilizer is most likely:

A. biased upward, because soil quality has a positive effect on yields and is left in the error term

B. unbiased, because soil quality is not included as a regressor

C. biased downward, since soil quality and fertilizer use are negatively correlated

D. biased, but the direction cannot be known without data on soil quality

[2 points] 13. In a regression of house prices on floor area, number of rooms and plot size, the auxiliary regression of floor area on the other two regressors has $R^2 = 0.90$. Compared with a case in which floor area was uncorrelated with the other regressors (and everything else was the same), the standard error of the coefficient on floor area is:

A. about 3.2 times as large, because the variance inflation factor is 10

B. about 10 times as large, because the variance inflation factor is 10

C. about 1.1 times as large, because the variance inflation factor is $\frac{1}{0.90}$

D. unchanged, but the coefficient on floor area is biased because the regressors are correlated

[2 points] 14. Suppose all the assumptions of the classical linear regression model hold except that the variance of the error term changes with the regressors (heteroskedasticity). The OLS coefficient estimates are then:

A. biased and inconsistent, so the model must be re-estimated by another method
B. unbiased and still the best linear unbiased estimators, so nothing needs to change
C. biased, but the bias disappears if heteroskedasticity-robust standard errors are used
D. unbiased, but the usual OLS standard errors, t-tests and F-tests are not valid

[2 points] 15. Which statement about heteroskedasticity-robust (White) standard errors is correct?

A. They change the OLS coefficient estimates as well as their standard errors
B. They are valid in large samples with or without heteroskedasticity
C. They are always larger than the usual OLS standard errors
D. They can be used only if the exact form of the heteroskedasticity is known

[2 points] 16. A researcher uses a single instrumental variable Z for an endogenous regressor X in the model $Y = \beta_0 + \beta_1 X + u$. Which of the two instrument conditions can be checked with the data?

A. Both: exogeneity by regressing Y on Z, and relevance by regressing X on Z
B. Exogeneity only, by checking that Z is uncorrelated with the OLS residuals
C. Relevance only, using the first-stage regression of X on Z
D. Neither, because both conditions involve the unobserved error u

[2 points] 17. In the first-stage regression of an endogenous regressor on a single instrument, the coefficient on the instrument is 0.42 with a standard error of 0.25. Which conclusion is correct?

A. The instrument is strong, because the first-stage F-statistic is greater than 1
B. The instrument is invalid: a weak first stage shows it is correlated with the error
C. The first stage is irrelevant for 2SLS provided the instrument is exogenous
D. The instrument is weak, since the first-stage F of about 2.8 is well below 10

[2 points] 18. A researcher has data on 300 firms, each observed for six years, and estimates a firm fixed-effects regression of log(output) on log(labour), log(capital) and a dummy variable family (1 if the firm is family-owned, a status that never changes during the six years). Which statement is correct?

A. Fixed effects removes the bias from every omitted variable, whether or not it changes over time
B. The coefficient on family cannot be estimated, because family does not vary within a firm
C. The fixed-effects estimator is identical to pooled OLS with one dummy variable for family ownership
D. Fixed effects is valid only if the firm effects are uncorrelated with labour and capital

[2 points] 19. A panel has two firms observed in two years. Firm P has $x = 2$ and $y = 10$ in year 1, and $x = 4$ and $y = 16$ in year 2. Firm Q has $x = 5$ and $y = 30$ in year 1, and $x = 7$ and $y = 34$ in year 2. The firm fixed-effects (within) estimate of the slope of y on x is:

A. 2.5
B. about 5.2
C. 3.0
D. 2.0

[2 points] 20. A researcher models whether a job seeker finds work within six months (job = 1 if so, 0 if not) using both a linear probability model (LPM) estimated by OLS and a logit model with the same regressors. Which statement is correct?

A. The LPM cannot be estimated by OLS, because the dependent variable takes only the values 0 and 1

B. Logit coefficients are marginal effects on the probability and can be compared directly with LPM coefficients

C. Logit predictions always lie between 0 and 1, but LPM predictions can fall outside that range

D. The LPM has homoskedastic errors by construction, so its usual standard errors are valid

Short Answer (4 questions)

[15 points] 21. Logs, dummy variables and interactions. A researcher uses a random sample of 1,500 full-time employees to estimate the earnings equation below by OLS. The dependent variable is log(wage), the natural log of hourly earnings in euros. educ is years of schooling, exper is years of work experience, female = 1 for women and 0 for men, union = 1 if the worker is covered by a collective agreement and 0 otherwise, and female × educ is the product of female and educ. Two cells of the output, [i] and [ii], are missing.

Regressor	Coefficient	Standard error	t-statistic
educ	0.0820	0.0060	13.67
exper	0.0150	0.0020	[i]
female	−0.3100	0.1240	−2.50
female × educ	0.0120	[ii]	1.50
union	0.1400	0.0350	4.00
constant	1.5500	0.0900	17.22
n	1,500		
R^2	0.312		

[2 points] (a) Find the missing values [i] and [ii].

[3 points] (b) Interpret the coefficient on educ, saying for which group it applies, and interpret the coefficient on union. For union, give both the approximate and the exact percentage effect.

[3 points] (c) What is the estimated return to a year of schooling for women? Test, at the 5 percent level, the null hypothesis that the return to schooling is the same for men and women against a two-sided alternative, stating the null in terms of a coefficient.

[3 points] (d) A classmate says that the coefficient on female, -0.310, shows that women earn about 31 percent less than comparable men. Explain why this reading is wrong, and estimate the gender difference in log wages for workers with 12 and with 16 years of schooling.

[2 points] (e) Construct a 95 percent confidence interval for the return to schooling for men. Can you reject, at the 5 percent level, the hypothesis that this return is 10 percent (a coefficient of 0.10)?

[2 points] (f) Interpret the R^2 of 0.312. Does the fact that about two-thirds of the variation is unexplained mean that the coefficient estimates are biased? Explain.

[15 points] 22. Functional form, F-tests and multicollinearity. An analyst uses data on 250 houses sold in one city to estimate the two models below by OLS. price is the sale price in euros, area is floor area in square metres, lot is plot size in square metres, bedrooms and baths are the numbers of bedrooms and bathrooms, and dist is the distance from the city centre in kilometres. Standard errors are in parentheses. With 244 degrees of freedom, the 5 percent two-sided critical value of the t distribution is 1.97.

Dependent variable: log(price)	(1)	(2)
log(area)	0.620	0.780
	(0.110)	(0.070)
log(lot)	0.180	0.190
	(0.045)	(0.045)
bedrooms	0.040	
	(0.028)	
baths	0.065	
	(0.040)	
dist	−0.025	−0.027
	(0.008)	(0.008)
constant	8.490	7.900
	(0.520)	(0.480)
n	250	250
R^2	0.640	0.625

[3 points] (a) Interpret the coefficients on log(area) and on dist in model (1).

[3 points] (b) In model (1), test at the 5 percent level the null hypothesis that the elasticity of price with respect to floor area equals 1, against a two-sided alternative. Explain what the result means.

[2 points] (c) Using two-sided tests at the 5 percent level, is each of bedrooms and baths individually significant in model (1)?

[4 points] (d) Test the joint null hypothesis that the coefficients on bedrooms and baths are both zero, at the 5 percent level. State the unrestricted and restricted models, calculate the F-statistic and compare it with the 5 percent critical value of the $F_{2,244}$ distribution, which is 3.03.

[3 points] (e) Explain how the results in (c) and (d) can both hold. What evidence in the table points to multicollinearity? Should the analyst drop bedrooms and baths from the model?

[15 points] 23. Omitted variable bias and instrumental variables. An economics department wants to know how much attending tutorials raises exam marks. For 600 students on a first-year course it records mark (the final exam mark, out of 100), attend (the number of the 10 weekly tutorials the student attended), gpa (grade point average in the previous semester, on a 0 to 4 scale) and early (= 1 if the student was assigned by lottery to a tutorial group that meets at 8 a.m., and 0 if assigned to an afternoon group; half the students were assigned to each). All four columns are estimated by OLS. Standard errors are in parentheses.

Dependent variable	(1) mark	(2) mark	(3) attend	(4) mark
attend	3.20	2.00		
	(0.20)	(0.23)		
gpa		10.00		
		(1.10)		
early			−2.00	−3.20
			(0.20)	(1.25)
constant	41.16	18.60	7.20	62.60
	(1.35)	(2.79)	(0.14)	(0.89)
n	600	600	600	600
R^2	0.300	0.384	0.143	0.011

[2 points] (a) Interpret the coefficient on attend in column (1). Give one reason why it is unlikely to measure the causal effect of attending a tutorial, and say in which direction it is likely to be biased.

[3 points] (b) A regression of gpa on attend gives a slope of 0.12. Show that the coefficients in columns (1) and (2) satisfy the omitted variable bias formula, and explain the sign of the bias. Is column (2) likely to measure the causal effect?

[3 points] (c) State the two conditions that early must satisfy to be a valid instrument for attend. For each condition, say whether it can be checked with the results in the table. Check it where you can, and otherwise discuss whether it is plausible here.

[4 points] (d) Calculate the instrumental variables (2SLS) estimate of the effect of attend on mark from columns (3) and (4). The software reports a 2SLS standard error of 0.55. Test whether attendance has an effect at the 5 percent level, give a 95 percent confidence interval, and compare the result with columns (1) and (2).

[3 points] (e) Give one reason why the exclusion restriction might fail in this setting, and explain in which direction it would bias the 2SLS estimate.

[15 points] 24. Binary outcomes and heteroskedasticity. Using survey data on 2,000 employed adults, a researcher studies whether a worker works from home at least one day a week (remote = 1 if so, 0 if not). inc is annual income in thousands of euros, degree = 1 if the worker has a university degree, and commute is the one-way distance to work in kilometres. Column (1) is a linear probability model (LPM) estimated by OLS, with the usual standard errors in parentheses and heteroskedasticity-robust standard errors in square brackets. Column (2) is a logit model, with standard errors in parentheses.

Dependent variable: remote	(1) LPM	(2) Logit
log(inc)	0.150	0.900
	(0.020) [0.024]	(0.150)
degree	0.120	0.720
	(0.025) [0.028]	(0.160)
commute	0.0040	0.0240
	(0.0018) [0.0022]	(0.0135)
constant	−0.450	−5.280
	(0.070) [0.085]	(0.550)
n	2,000	2,000
R^2	0.094	0.082 (pseudo)

[3 points] (a) Interpret the LPM coefficients on degree and on log(inc). For log(inc), give the predicted change in the probability of remote working when income rises by 10 percent.

[3 points] (b) Are the usual (non-robust) OLS standard errors valid in an LPM? Explain. Using the appropriate standard errors, test at the 5 percent level, against a two-sided alternative, whether commuting distance affects the probability of remote working. Would the usual standard errors lead to a different conclusion?

[3 points] (c) To test for heteroskedasticity formally, the researcher regresses the squared LPM residuals on log(inc), degree and commute; this auxiliary regression has $R^2 = 0.021$. Carry out the Breusch-Pagan LM test at the 5 percent level. The 5 percent critical value of the χ^2_3 distribution is 7.81.

[3 points] (d) Use the LPM to predict the probability of remote working for a worker earning 10 thousand euros a year, with no degree, who commutes 5 km. Then do the same with the logit model. Comment on the two answers.

[3 points] (e) For a worker with a degree, an income of 40 thousand euros and a 20 km commute, show that the logit model predicts a probability of about 0.32. Calculate the logit marginal effect of log(inc) for this worker and compare it with the LPM coefficient. Why are the logit coefficients so much larger than the LPM coefficients?



Make one like this from your own notes.

Scan the code or type the address below: upload your lecture notes, slides or a past paper and get a practice exam with an answer key in about a minute. Free, no card.

`examtex.com/create`