Foundations of Econometrics - Part I

Final exam: sample

Please use this sample of exam questions as an indication of the style of questions you might encounter in your exam, but not as an indication of the content of the exam. That is, topics covered during the course that do not appear below might be included in your exam.

1. *(7 points)* Consider the following model relating a country $GDPPC$ (US$), in log form, with dem=1 if country is defined as a democracy, 0 otherwise, $educP$=% of country's population with primary schooling attained, $educS$=% of country's population with secondary schooling attained , $educH$= % of country's population with higher education attained , and $tradewb$=Exports plus Imports as a share of GDP.

$$Model(1): \quad ln(GDPPC) = \beta_1 + \beta_2 dem + \beta_3 educP + \beta_4 educS + \beta_5 educH + \beta_6 tradewb + \epsilon.$$

This regression model is estimated using data for 2010 from Acemoglu's article, "Democracy does cause growth". *Stata*'s output after estimating by OLS, is:

Source	SS	df	MS	Number of obs	=	122
				F(5, 116)	=	34.46
Model	182.578494	5	36.5156988	Prob > F	=	0.0000
Residual	122.910661	116	1.05957466	R-squared	=	0.5977
				Adj R-squared	=	0.5803
Total	305.489155	121	2.52470376	Root MSE	=	1.0294

lnGDPPC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dem	.380598	.2183563	1.74	0.084	-.0518841	.81308
educP	.0267492	.0073504	3.64	0.000	.0121908	.0413076
educS	.0300333	.0061537	4.88	0.000	.0178452	.0422215
educH	.0859244	.0098279	8.74	0.000	.066459	.1053898
tradewb	.004786	.0020264	2.36	0.020	.0007724	.0087996
_cons	4.094374	.4496879	9.10	0.000	3.20371	4.985037

(a) *(1 point)* (i) Provide the general expression that has been used to calculate the standard error, value 0.2183563 underlined in the output above, in as much detail as possible. Then, substitute the available values provided in the output. Any element missing? If so, properly identify it. (ii) What information is this value meant to provide? Be as specific as you can.

(b) *(1 point)* Consider testing whether regressor dem is statistically significant using the exact t-test statistic.

(i) Indicate null and alternative hypotheses.

(ii) For this test, indicate the expression of the test statistic and its assumed distribution.

(iii) List by name all the assumptions we would need to place on the dgp for the t-test statistic to have the distribution you indicated in (ii).

(iv) At what significance level would regressor dem be statistically significant? Justify.

(c) *(0.5 points)* (i) Provide the exact expression that defines the $p-value = 0.084$, underlined in the output above. (ii) Draw this $p-value$, properly identifying both axes and any relevant value to properly identify this $p-value$

(d) *(1.5 points)* Consider testing whether $educP$, $educS$ and $educH$ are jointly significant at $\alpha = 1\%$, using the exact F test statistic with the following assumed distribution:

$$F \equiv \frac{(RSSE - SSE)/q}{SSE/(n-K)} \underset{\text{under} H_o}{\sim} F(q, n-K).$$

(*i*) Write down the null and alternative hypotheses associated with this test.

(*ii*) Using the information provided below about the $F-distribution$, draw the acceptance and rejection region associated with this test. Properly label the axes. Include any relevant value to properly identify both regions.

(*iii*) Provide the intuition behind where the acceptance region is located.

(*iv*) Provide the exact expression of the model you would have to estimate to calculate $RSSE$.

(*vi*) Knowing that the associated $F-value = 41.3$, what would you conclude? Justify.

Useful information about a $F \sim F(3,116)$ *distribution*:

$F \sim F(3,116)$: $Prob\{F > 2.23\} = 0.1$, $Prob\{F > 2.68\} = 0.05$, $Prob\{F > 3.96\} = 0.01$.

(e) *(1 point)* Consider again the test included in question (1b). We wanted to drop classical assumption of disturbances being conditionally normal. Without normality assumption,

(*i*) Would the distribution of the statistic you used in (1b) change? How?

(*ii*) Would the acceptance region you drew in (1b)(*iii*) be affected? Can you tell how? Explain.

(f) *(1 point)* If we wanted to drop classical assumption of disturbances being homoskedastic, but still wanted to use OLS estimator to estimate model parameters,

(*i*) Would it affect how R^2 has to be calculated? Justify.

(*ii*) Would it affect how the standard errors have to be calculated? Justify.

(g) *(1 point)* Acemoglu et al.(2019)'s article "Democracy does cause growth", included the following paragraph: *"The estimation of the causal effect of democracy on GDP faces several challenges. Democracies differ from non democracies in unobserved characteristics that can also have an impact on GDP. As a result, cross-country regressions are unlikely to reveal the causal effect of democracy on growth."* Rigorously discuss Acemoglu et al.'s argument using the regression presented in this question.

2. *(3 points)* Consider the following *dgp*:

$$\begin{aligned} y_i &= 2 + 0.5 \cdot x_{i2} + 0.5 \cdot x_{i3} + \epsilon_i \quad \epsilon_i / X \sim i.i.N(0,16) \\ & x_{i2} \sim U[0,20] \qquad x_{i3} = x_{i2} + v_i \quad v_i \sim N(0,\sigma_v^2) \end{aligned}$$

After setting $\sigma_v^2 = 9$ we use this *dgp* to generate 10,000 samples of 30 observations each ($n = 30$). With each of the generated samples, we estimate by OLS the following regression:

$$(1) \qquad y = \beta_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \epsilon,$$

and save the estimates of parameter β_2. Paying careful attention to the *dgp* and the regression we estimate with each generated sample, answer the following questions:

(a) *(1 point)* Consider we summarize the 10,000 OLS estimates of β_2 we got using a density histogram. How would the graph, approximately, look like (shape, location)? Draw the histogram, label both axes, and justify both, the shape and location you selected.

(b) *(1 point)* Now, we repeat the experiment, but setting $\sigma_v^2 = 1$. How would how would you expect the histogram you drew in question 2a to change (location, spread, shape)? Rigorously argue using the expression you feel is most appropriate to support your argument.

(c) *(1 point)* Consider that using the same samples generated in 2a we estimate instead regression:

$$(2) \qquad y = \beta_1 + \beta_2 \cdot x_2 + \epsilon.$$

Would you expect the location of the histogram to change? If you do not, explain why. If you do, explain why and how. Rigorously justify in either case.

3. *Extra question(1 point)* For a linear regression model with 4 regressors $(const, x_2, x_3, x_4)$, we want to test: $H_0 : \beta_2 = \beta_3$ versus $H_1 : not\ H_0$, using an estimator *tilde*, $\tilde{\beta}$, (i.e., not the OLS estimator!), which has the following asymptotic distribution:

$$\sqrt{n}(\tilde{\beta} - \beta) \overset{a}{\sim} N(0, \Omega)$$

   (a) Specify the dimension of the following elements of the asymptotic distribution above: $\tilde{\beta}$, β, 0, Ω. No need to justify.

   (b) Derive, step by step, a test statistic to perform the test above using this estimator tilde, for the case where matrix Ω is known. Justify why the statistic you derived is a proper test statistic.