Assessment guidelines, exam SØK3001 Advanced Econometrics, fall 2023

This is guidelines for assessment. Thus, it is not a complete suggestion of solution. The presentation here is shorter than expected for a complete solution.

Question 1

Briefly explain the following concepts:

- Measurement error

When an econometric model uses an imprecise measure of the variable that belongs to the model, then the model contains measurement error. The observed variable is not identical to the variable that belongs to the model. It might be useful to distinguish between measurement error of explanatory variable(s) and the dependent variable.

- Dummy variable

A dummy variable takes only two values, typically the values 0 and 1. Thus, it is sometimes denoted a qualitative variable.

- Heteroscedasticity

The variance of the error term of the model, conditional on the explanatory variables, is not constant.

- Instrumental variable

Instrumental variables are relevant in models with endogenous explanatory variable(s). An instrumental variable is correlated with the endogenous variable and is uncorrelated with the error term of the model.

Question 2

A researcher studies how the interest rate and income affect housing prices. Using 80 quarterly observations, the researcher reports the following estimated housing price equations:

(1) $ph_t = \underset{(0.030)}{-0.060} - \underset{(0.040)}{0.005}\, r_t - \underset{(0.002)}{0.015}\, r_{t-1} + \underset{(0.035)}{0.350}\, y_t + \underset{(0.235)}{0.950}\, y_{t-1}$

(2) $ph_t = \underset{(0.030)}{-0.020} + \underset{(0.020)}{0.80}\, ph_{t-1} - \underset{(0.0025)}{0.008}\, r_t + \underset{(0.035)}{0.330}\, y_t$

(3) $ph_t = \underset{(0.030)}{-0.020} + \underset{(0.025)}{0.70}\, ph_{t-1} - \underset{(0.045)}{0.002}\, r_t - \underset{(0.003)}{0.012}\, r_{t-1} + \underset{(0.040)}{0.120}\, y_t + \underset{(0.080)}{0.440}\, y_{t-1}$

where ph is the log of housing prices, r is the interest rate in per cent and y is the log of real disposable income. Standard errors are reported in parentheses below the estimated parameters. The estimation method is OLS.

a) Give an interpretation of the results reported in equation (1) – (3). Find the short- and long-run effects of the interest rate and income on housing prices and discuss how fast housing prices adjust to changes in the explanatory variables.

Notice that ph and y are in logs so the effects of income discussed below are elasticities. Since r is untransformed, the effects of the interest rate are interpreted as semi-elasticities: Multiply by 100 we obtain the change in house prices in per cent if the interest rate increases by one percentage point.

Eq (1). The short-run effect of interest rate is -0.005; the long-run effect is -0.005 – 0.015 = -0.02. The short-run effect of income is 0.35; long run effect is 0.35+0.95=1.3. Given this specification, house price is completely adjusted one quarter after a change in the exogenous variable. Should notice that the delayed effect is much higher than the immediate effect for both variables.

Eq(2): The short-run effect of interest rate is -0.008 and the short-run effect of income is 0.33 (both close to the SR-effects in (1)). The candidate should explain what is meant by long-run effects in an equation with lagged endogenous variable: Assume that the explanatory variables increase permanently. Assume further that ph converges to a new equilibrium level after a permanent increase in r and y. To calculate the long-run (steady state) effects set $ph_t = ph_{t-1} = ph^*$ *in (2) and solve for ph*, which gives:*
ph =-r* 0.008/0.2 + y*0.33/0.2 + const = -0.04r +1.65y.*

The parameter in front of ph_{t-1} *= 0.8 indicates relatively high degree of persistence, or sluggish adjustment (can be further discussed under b)).*

Equation (3) is a more general specification where (2) is expanded with lags of r and y. The short-run effect of r = -0.002, of income 0.12. To find the long run effects, assume again that r and y increase permanently and find the new steady state effect on ph: For r = -(0.002+0.012)/0.3 = -0.047; for y = (0.12+0.56)/0.3 = 1.87. Same interpretation of the estimate in front of lagged house price as in (3).

The researcher also estimates a fourth specification and reports the results given by:

(4) $\Delta ph_t = -\underset{(0.030)}{0.020} - \underset{(0.015)}{0.30}\, ph_{t-1} - \underset{(0.045)}{0.002}\, \Delta r_t - \underset{(0.003)}{0.014}\, r_{t-1} + \underset{(0.040)}{0.120}\, \Delta y_t + \underset{(0.090)}{0.560}\, y_{t-1}$

b) Use the results in equation (4) to find the short- and long-run effects of the interest rate and income. Compare the effects with those derived from equation (3).

Should notice that (4) is a transformation of (3) so we should obtain the same effects (both SR and LR). We see immediately that the SR-effects are equal. To calculate the LR-effects, assume that the delta terms in (4) are all constant in a new steady state (not necessarily = 0). Then solve for ph and obtain the long-run effect of r equal to-0.014/0.3 = -0.047, of y = 0.56/0.3 =1.87.

c) Discuss how the estimated parameter in front of ph_{t-1} can be interpreted.

The absolute value of the parameter in front of ph_{t-1} *in (4) is interpreted as the "speed of adjustment parameter" and tell us that 30% of a deviation from the new steady state is eliminated during one* *quarter. Could notice that the parameter in front of ph_{t-1} in (4) is equal to the parameter in front of ph_{t-1} in (3) minus one: -0.3 = 0.7 – 1 (consistent with the transformation).*

d) Housing prices can be expected to show systematic seasonal variations. Explain how this property can be taken into account in a revised housing price equation.

Simply include 3 quarterly dummy variables, for example one for Q1, one for Q2, and one for Q3 if the constant term is included.

The empirical study may be problematic since the included variables are non-stationary.

e) Explain what is meant by a non-stationary variable and further explain how you can test statistically whether a variable is non-stationary.

A non-stationary variable does not have constant mean, constant variance and constant autocorrelation (the requirement for a stationary variable)

Test: Explain the Dickey-Fuller test using an equation like:

$\Delta x_t = \rho * x_{t-1}$ *+ constant.*

The null hypothesis is $\rho=0$. The alternative is $\rho < 0$. Estimate ρ, calculate the t-statistic and compare with critical values based on the Dickey-Fuller distribution. In general, if the t-statistic is sufficiently below zero, we reject the null hypothesis that x is non-stationary.

f) Explain how you can test statistically whether ph cointegrates with r and y.

Explain the Engle-Granger 2-step procedure: First, estimate a static regression with ph on the left hand side, and r and y on the right. Second, find the residuals from this regression and use the Dickey-Fuller test described under e) to test whether the residual is non-stationary. If we reject the null hypothesis, and conclude that the residual is a stationary variable, we say that ph cointegrates with r and y.

g) Assume that ph cointegrates with r and y. Explain how we can interpret such a relationship between the variables.

If ph cointegrates with r and y, the estimated static regression in f) is interpreted as a long-run equilibrium equation. The residuals are interpreted as deviations from the equilibrium relationship, and if the residuals are statistically, these deviations will converge towards zero after a shock.

Question 3

Higher wages might be a motivation for higher education. The average effect of years of education on the wage is therefore of interest. For the age group 28-38 years, we have access to a random sample of workers. The data include the monthly wage (Wage), years of education (Education), years of job experience (Experience), whether the worker lives in a city (Urban), and the result of an IQ test (IQ). We are interested in variations of the following model:

$$\log(wage)_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_{ii} + \beta_3 Urban_i + \beta_4 IQ + u_i$$

Where the lower letter i denotes the individual and u is the residual.

a) What are the necessary assumptions for unbiased estimates of the coefficients and the standard errors when using the least square method (BLUE)?

The assumptions for unbiased coefficients are (i) linearity in parameters, (ii) random sampling, (iii) no perfect collinearity, and (iv) zero conditional mean. In reality, the last assumption is the most challenging in economic analyses. It is expected that some explanations are provided for the assumptions, in particular (ii) and (iv). The assumptions for unbiased standard errors include additionally (v) constant variance of the error term (no heteroskedasticity).

b) Describe the consequences of an omitted variable in the model.

The consequences depend on the correlation with the included explanatory variables. If there is no correlation, the estimated coefficients are unbiased. If there is correlation, the estimates will be biased if there is an effect of the omitted variable. It is useful to give a formal presentation. One possibility is to use the model above, but only including two variables. That is, for example, the model $\log(wage)_i = \beta_0 + \beta_1 Education_i + \beta_2 Experience_i + u_i$.

The question is the effect of omitting, say, Experience. It can be shown that the estimate of β_1 is given by $E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \tilde{\delta}$*, where* $E\left(\tilde{\beta}_1\right)$ *is the expected coefficient in the case where IQ is excluded from the model and* $\tilde{\delta}$ *is the coefficient of Education when regressing Experience on Education. The size of the bias depends on the size of the correlation between Experience and Education (measured by the regression coefficient). If β_2 is positive, the bias is positive (overestimating β_1) if the correlation is positive and the bias is negative if the correlation is negative.*

Results for different variants of the model are presented in Table 1. The table presents estimated coefficients, with standard errors in parentheses.

c) What is the economic interpretation of the estimated coefficient of Education in column (1)?

When Education increases by one year, the wage increases by 0.06 log points. That is about 6 percent.

d) Do a hypothesis test for whether the coefficient is equal to zero.

It is expected that the hypothesis test is formally presented, where H_0 is $\beta_1 = 0$. Then the t-value is calculated as $t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ *with n-k-1 degrees of freedom. It follows that t = 10 with 933 degrees of freedom. Using the table for critical values, it must be concluded that the effect is significantly different from zero at whichever significance level chosen.*

$t = \hat{\beta}_1$

e) The estimated effect of Education changes from column (1) to column (2), and from column (2) to column (3). Explain why.

The changes follow from more variables included. This implies that the model in both column (1) and (2) has an omitted variable bias. Notice that all coefficients in the table are positive. Comparing column (2) and (3), it can be concluded, by referring to the answer on b), that Education and IQ is positively correlated and that IQ and Experience and IQ and Urban are uncorrelated. Comparing column (1) and (2), it can be concluded that either Experience or Urban is negatively correlated with Education. One mechanism might be that those taking higher education have been working fewer years than those without higher education, which implies a negative correlation between Education and Experience.

f) The variable IQ has a mean value of 100 and a standard deviation of 15. Based on the estimated model, how much higher expected wage has an individual with two standard deviations higher IQ than the average, all else equal?

*Two standard deviations are a change of 30. Given the marginal effect of 0.006, the change of 30 is equal to 30*0.006 = 0.18. The wage is expected to be about 18 percent higher when IQ increases by 2 standard deviations.*

g) A commentator claims that the choice of higher education to a large extent depends on parental background and their influence during childhood. Discuss different econometric approaches that can be used if this claim is correct.

One might see this as an omitted variable problem, which can be solved by including variables for parental background. It is expected that it is discussed whether "influence during childhood" is measurable. If not, the challenge cannot be solved by including more variables. The empirical approach must take this into account. There are two possible approaches on the reading list. It is not expected that the answer goes into how realistic the approaches are in real analyses. The difference-in-differences approach might be used when there is an external reform that increases Education from one year to another. The instrumental variable approach can be used if a valid instrumental variable exists.

h) Another commentator makes you aware that the data used in the analysis are from the USA. Do you think the results are valid also for Norway? Explain.

There is no final answer on this question. It is about the external validity of the results. On might argue that the educational and labour market institutions differ to a large extent between the countries, and thus question the external validity.

Tabell 1. Estimation results

	(1)	(2)	(3)
Education	0.060 (0.006)	0.076 (0.006)	0.055 (0.007)
Experience	-	0.020 (0.003)	0.020 (0.003)
Urban	-	0.173 (0.028)	0.173 (0.028)
IQ	-	-	0.006 (0.001)
Konstant	5.973 (0.081)	5.405 (0.111)	5.101 (0.120)
R-kvadrert	0.10	0.17	0.20
Observasjoner	935	935	935