SØK 3001 H 22. Examination guidelines

Question 1.

a) Exogeneity/endogeneity in econometric models is discussed in Woolridge chp. 3.3.

Let the econometric model be

(1) $y=\alpha+\beta x+u$

Briefly explained, exogeneity requires the assumption E(u|x)=0, which implies that cov(x,u)=0 and E(u)=0, i.e. the explanatory variable is uncorrelated with the error term. Endogeneity is that E(u|x) ≠0 which implies cov(x,u) ≠0. Reasons: Omitted variables, measurement error in x, or that x and y er simultaneously determined. Some explanations of each of these sources should be included, but extended and long explanations are not required. Can briefly explain that OLS estimator for β in the equation (see chp. 5 in Woolridge) in the absence of exogeneity becomes inconsistent i.e. $plim\ \hat{\beta} \neq \beta$ where $\hat{\beta}$ is OLS estimator for the unknown population parameter β. Should point out that the OLS-estimator for β is inconsistent estimator if x is not exogeneous (i.e. u is correlated with x in (1)).

b) An instrumental variable is a variable z that is correlated with the explanatory variable x, but uncorrelated with the error term u in (1), more precisely the requirements are that cov(z,u)=0 (exclusion requirement) and cov(x,z)≠ 0 (relevance requirement). If these requirements are fulfilled, the IV-estimator $\hat{\beta}^{iv}$ is a consistent estimator in terms of $plim\ \hat{\beta}^{iv} = \beta$. Variables z1; z1 and z2; and z1, z2 and z3 are used as instruments in the estimation in columns (2), (3) og (4) in Table 1, respectively.

IV-estimators in columns (2), (3) og (4) are all consistent i.e. plim $\hat{\beta}^{iv}$ =β **if** z1; z1 and z2; and z1, z2, z3 are valid instrumental variables for x, (see chp. 15 in woolridge), formally

cov(z1, u)=0, cov(z2, u)=0 and cov(z3,u)=0 (exclusion requirements)

cov (x, z1) ≠0, cov(x, z2) ≠0 og cov (x, z3) ≠0 (relevance requirements)

c)Estimates in columns (2), (3) og (4) are 2SLS estimates, where 1.stage is the OLS regressions in columns (5), (6) og (7), respectively

(5) $x = \pi_0 + \pi_1 z1 + error$, (6) $x = \pi_0 + \pi_1 z1 + \pi_2 z2 + error$ and

(7) $x = \pi_0 + \pi_1 z1 + \pi_2 z2 + \pi_3 z3 + error$

Estimated coefficients in first stage equations (5), (6) and (7) are used to predict x, $\hat{x}$. $\hat{x}$ can then be used as instrumental variable for x in the structural equation. The method can be implemented by replacing x with $\hat{x}$, in the structural equation (1) and estimating the resulting equation with OLS in the second stage, see chp. 15 in Woolridge.

Testing relevance of z1; z1 and z2; z1, z2 and z3, respectively, as instruments in the equations estimated in columns (2), (3), (4) imply testing if z1; z1 and z2; z1, z2 and z3 contributes significantly to explain the variation in x represented by the l.stage-equations estimated in columns (5), (6) and (7), respectively in Table 1. Example: Test

H0: $\pi_1 = \pi_2 = 0$, in column (6)

Can use the R-squared version av F-test that all explanatory variables in an OLS-regresssion contribute to explain x (see Woolridge chp. 4-5).

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Here k=2, n=100 and R^2 =0. 333 in column (6) in Table 1, which gives F=24.2 and the null hypothesis is clearly rejected, suggesting that instruments are relevant. F-value clearly exceed the rule of thumb border of F=10. Similar F-test can be performed for the case with 3 instruments in column (3) with F-test from column (7). The case with one single instrument as in col (2) implies a simple t-test on the effect of z1 in col (5).

d) Exogeneity test is explained in Woolridge chp. 15 p. 481, and implemented by estimating OLS-regression as reported in col (8) where the structural equation is extended with the residual, e5, from 1.stage in 2SLS regression (col (5)). Intuition is that if x is exogeneous and z1 is a valid instrument, the residual e5 which represents the variation in x that don't come from z1 should not contribute significantly to the explanation of y. Thus, exogeneous x implies the hypothesis that the cofficient in front of e5 in col (8) equals zero. t-statistic at 0.714/0.187 implies that the null hypothesis of exogeneous x is rejected.

e) Test of overidentification is considered in Woolridge chp. 15 p. 482. Identification in an equation with one endogeneous right hand side variable implies that we have *at least one* exogeneous variable (one instrument) which is correlated with x but uncorrelated with the error term in the structural equation (1). Exact identification implies that we have *exactly one* valid instrument, while overidentification implies that we have *more than one* valid instrument. Intuition is that if for instance z1 and z2 are both valid instruments, the IV-estimators based on respectively z1 and z2 as instruments, are both consistent and therefore not systematically different. As explained in Woolridge p. 483-484 the over-identification test can be implemented by estimating an auxiliary regression with the residual from for instance the 2SLS regression, e3, as dependent variable with z1 and z2 as regressors and create the test statistic $n \cdot R^2$, where R^2 is the R-squared from the auxiliary regression. Under H0 of overidentification, the test statistic is Chi-square distributed with degrees of freedom equal to the number of overidentification restrictions which here equals 1. In col (9) the test statistic becomes 0.363·100=3.63 which is lower than the critical value at 5% significance level (3.84). Thus, we cannot reject the hypothesis of overidentification in the case with z1 and z2 as instruments. On the other hand, we must reject the null hypothesis of overidentification in the case with z1, z2 and z3 as instruments since the test statistic in this case becomes 0.131·100=13.1, exceeding the critical value in the chi-square distribution. These results may suggest that z3 is a suspect instrument potentially correlated with the error term u in (1).

Question 2.

a)Here is it useful at the start to explain that the model formulation (1) with current values of both y, p and pb imply that changes in the price of electricity in a period changes the use of electricity in the same period (momentaneous adjustment). The log-log formulation implies that the parameters can be interpreted as the own price elasticity and cross-price elasticity of the use of electric power.

b) and c) Compared to (1) additional investments required to change the use of electricity can be interpreted as a situation where changes in p affects y with time lags. Here is it relevant to formulate different model specifications explicitly. Here is an example with an ADL-model which is relevant to represent dynamics.

(1)' $y_t = \beta_0 + \gamma y_{t-1} + \beta_{10} p_t + \beta_{11} p_{t-1} + \beta_2 pb_t + u_t$

The ADL-model (1)' can be re-parameterized as an error correction model that is useful to apply to test hypothesis about short run and long run electricity price elasticities in the next question d).

(1)'' $y_t - y_{t-1} = \beta_0 + (\gamma - 1) y_{t-1} + (\beta_{10} + \beta_{11}) p_{t-1} + \beta_{10}(p_t - p_{t-1}) + \beta_2 pb + u_t$

$(\beta_{10} + \beta_{11})/(1-\gamma)$ is longterm price elasticity and β_{10} is short-term price elasticity d)Here, the statement of the commentator can be interpreted as statements of the size of the short-run and long-run price elasticities for electricity and the candidates should present test-procedures based on different possible interpretations of the statement.

The statement **may** be interpreted as a statement that the short run price elasticity equals zero, H01: $\beta_{10} = 0$, while the long-run elasticity is allowed to be < 0. This can be tested by a simple t-test of the significance of β_{10} based on OLS on (1)' or (1)''

Alternatively the statement **may** be interpreted as a statement that **both** the short-run **and** long run price elasticities are equal to zero, which implies

H02: $\beta_{10} = 0$ **and** $\frac{\beta_0+\beta_1}{1-\gamma} = 0$ which in (1)'' implies that H02:$\beta_{10} = 0$ **and** $\beta_{10} + \beta_{11} = 0$

H02 can be tested with an F-test of the hypothesis that the coefficients in front of $p_t - p_{t-1}$ and p_{t-1} in (1)'' are simultaneously equal to zero.

Question 3.

a) logarithmic version of Cobb Douglas is $ly = \beta_0 + \beta_1 ll + \beta_2 lc + u$

where β_1 is the marginal output elasticity with respect to employment and β_2 the marginal output elasticity with respect to capital and u is a stochastic error term.

b) Columns (1) and (2) are based on pure cross-section data and without variables that represent technology and management quality, the effect of such variables cannot be accounted for in these regressions and we may possibly have an omitted variable problem to the degree that such variables are correlated with the included ll and lc. The OLS estimators for β_1 og β_2 will thus potentially be biased and inconsistent.

The estimates reported in Columns (3)-(4) i.e. pooled OLS will also potentially be biased and inconsistent for the same reason because they don't account for the impact of neither permanent management quality differences nor common technical change.

Columns (5) and (6) are pooled OLS with year dummies that account for the effect of common technological change affecting all firms in China, but these regressions do not account for the effects of management quality.

Column (7) and (8) is OLS with fixed firm effects accounting for the effect of permanent mangement quality differences across firms, but since year dummies are not included, these regressions does not account for common technological change across firms.

Columns (9) and (10) are OLS with fixed firm effects and year dummies and account for both permanent differences in management quality and common technological change

b) Restrictions:

Columns (1) and (2): All coefficients (including the constant term) are equal across firms, and the effects of capital and employment are equal across firms

Columns (3) and (4): Coefficients are equal in all time periods (i.e constant term and the effects of capital and employment do not vary across firms and years).

Columns (5) and (6): All coefficients are equal across firms. The constant term are constant across firms, but varies across years and account for year specific differences across years in technological progress common for all firms.

Columns (7) and (8): All coefficients except the constant term are equal across firm. The constant term varies across firms and account for permanent differences in the production function (including management quality) that are equal across firms over time.

Columns (9) and (10): All coefficients, except the constant term are constant across firms and the constant term varies between firms and across years.

c) and d) Here is it important to realize that the models in col (2) is a pure reparameterization of the models in col (1), (4) is a reparameterization of (3), (6) is a reparameterization of (5) etc. This is important because it allows a simple way of testing the hypothesis of constant returns to scale in question d). The coefficient in front of ll in columns (2), (4), (6), (8) and (10) equals 1 minus the sum of marginal elasticities of output with respect to employment and capital, 1-$(\beta_1 + \beta_2)$, while the coefficients in front of (lc-ll) in (2), (4), (6), (8) and (10) is the marginal elasticity of output with respect to capital (β_2). The estimated marginal elasticities of employment and capital are thus 0.271 and 0.074 in columns (9) and (10.

95% confidence interval is approximately equal to

coefficient estimate$\pm 2se$,

where se=estimated standard error of the coefficients

Constant returns to scale imply the hypotheses H0: $(\beta_1 + \beta_2) = 1$, Testing implies a twosided test on the hypothesis that the coefficient $1 - (\beta_1 + \beta_2)$ in front of ll in respectively columns (2), (4), (6), (8) and (10) equal zero. Interpretation of the hypothesis: Constant returns to scale implies that p% increase in both inputs increase output by p%. t-values should be computed and commented.

Good candidates should connect the choice of credible specifications to the differences between them in terms of potential omitted variables, cf. question a). Based on this view, results in columns (9) and (10) provide the most credible results since they account for both permanent differences across firms as well as the impact of common technical progress across all firms in China.

e) Here the simple answer is that since PRIVATE as defined does not vary within firms, the variable will be perfectly correlated with the firm dummies and the effect of this variable cannot be estimated in specifications including firm specific dummies.

f) Easy to extend the models with firm fixed effects with an interaction term $lc \cdot PRIVATE$. Since this variable varies both in the cross section and time dimension, it is possible to identify the effect of the variable also in the versions with firm specific dummies included. If the coefficient in front of $lc \cdot PRIVATE$ is positive (negative) it indicates that the marginal elasticity of output with respect to capital is higher (lower) in privately owned firms than in publicly owned firms.

g) It is possible to include such a variable (WP) in the specifications without year dummies. Good candidates would realize and comment on the fact that consistent estimates of the effect of WP in these specifications requires that this variable does not account for the effects of other variables that also only varies over time (macro variables). WP will be perfectly correlated with year dummies and cannot be included in model specifications where these are included.