Exam in: ECON 4160: Econometric Modelling and System Estimation

Day of exam: 25 May, 2009

Time of day: 9:00-12:00

This is a 3 hour school exam.

Guidelines:
Try to answer all questions.
For reference, some relevant critcal values for the F distribution is given at the end of the question set.

1. Assume that y_1, y_2, ...y_n are n stochastic variables. Assume that, based on economic theory, a hypothesis is formulated saying that "y depends on x".

   (a) Explain how this hypothesis can be tested with the use of equation

   $$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i,\ i = 1, 2, 3, ..., n, \quad (1)$$

   and ordinary least squares (OLS) estimation. (Note: take care to define the symbols in (1) as part of your answer).

   Formulate H_0: $\beta_2 = 0$ and test using the t-distribution.

   (b) Does it matter for you answer to 1a) whether the explanatory variable x_i is deterministic or stochastic? Explain briefly.

   No it does not matter. If y_i and x_i have a joint pdf then there is a conditional distribution for y_i given x_i, and the conditional expectation of that PDF is the regression function. If $f(x_i, y_i)$ $i = 1, 2, 3, ..., n$ are IID, we can define disturbances as $\varepsilon_i = y_t - E[y_i \,|x_i]$, that have classical properties, and if the joint pdf is normal, ε_i is also normal. Equivalently, and alternatively, refer to iterated expectations and that OLS is unbiased both conditionally and unconditionally, and that t-statistics will give correct inference also with stochastic x_i.

   (c) Assume that a fellow student remarks critically that inference based on OLS estimation of (1) is misleading, because, as she says: "the disturbances in equation (1) may be heteroscedastic". How would you respond to this critique?

   In the case of het, OLS still gives unbiased and consistent estimators, but the BLUE property is lost. This gives rise to misleading inference. Two pther points: If the form of het is known, use GLS. If the form of het is unknown, use White's het consistent standard errors to robustify inference.

2. Assume that we have a sample of observations of total income (not total consumption expenditure) for n households. The income variable is denoted x_i ($i = 1, 2, .., n$). We also have observations of J commodity expenditures for the

n households. We denote the expenditures by y_{ji} ($j = 1, 2, ...J$; $i = 1, 2, ...n$). We want to estimate the parameters $(\alpha_{j0}, \alpha_{j1})$ in the J equations:

$$\begin{aligned} y_{1i} &= \alpha_{10} + \alpha_{11} x_i + \varepsilon_{1i},\ i = 1, 2, ..n. \\ y_{2i} &= \alpha_{20} + \alpha_{21} x_i + \varepsilon_{2i},\ i = 1, 2, ..n. \\ &\vdots \\ y_{Ji} &= \alpha_{J0} + \alpha_{J1} x_i + \varepsilon_{Ji},\ i = 1, 2, ..n. \end{aligned} \tag{2}$$

where $\varepsilon_{1i}, ..., \varepsilon_{Ji}$ denote the disturbance terms.

(a) Assume that the disturbances in each equation of (2) are independent and identically distributed, but that there are correlations between the disturbances in the different equations. Give one or more reasons for why this kind of correlation is relevant for the "expenditure system" in (2).

Individual effects that affect demand for all commodities. Budget constraint.

(b) Would you choose the SURE or the OLS estimator to estimate the parameters $(\alpha_{j0}, \alpha_{j1})$? Explain you anwer.

Generally SURE is efficent and unbiased if the disturbances are correlated. However if the same explantory variable appears in all equations the SURE estimator and the OLS estimator give identical results. This is the case here.

(c) Assume that x_i is an unobservable variable. Instead of x_i, we observe x_i^* which is defined by the equation

$$x_i^* = x_i + v_i, \tag{3}$$

where v_i is a stochastic variable that satisfies the classical disturbance assumptions, and is independent of both ε_{ji} and x_i. How would this change in model specification affect your choice of estimation method?

By assumption v_i is correlated with x_i^*, so measurement errror with consequences.

3. Explain how you would estimate the parameters *in each equation* of the following model.

$$\begin{aligned} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} + \gamma_{12} x_{2t} &= \varepsilon_{1t} \\ y_{2t} + \gamma_{21} + \gamma_{23} x_{3t} &= \varepsilon_{2t} \\ \beta_{32} y_{2t} + y_{3t} + \gamma_{31} + \gamma_{33} x_{3t} &= \varepsilon_{3t} \end{aligned} \tag{4}$$

where y_{1t}, y_{2t} and y_{3t} are endogenous variables.

HINT: Take care to base your answer on a complete econometric specification of the model.

If ε_{it} $(i = 1, 2, 3)$ cannot be assumed independent: The first equation is just identified. 2SLS, IV and ILS are all equivalent and provide consistent estimators. x_3 is used as instrument for y_2. The second equation contains no *endogenous explanatory variables, hence OLS of y_{2t} on y_{3t} will provide unbiased and consistent estimators. The third equation is not identified.*

If ε_{it} $(i = 1, 2, 3)$ can be assumed independent: After re-ordering the system is recursive. OLS on each equation will provide unbiased and consistent estimators

4. In this question we discuss the empirical relationship between the money market interest rate, denoted R_t, and the banks' loan interest rate in the Norwegian Central Bank, denoted CBR_t. The subscript t denotes time period. The initial hypothesis is that in an inflation targeting regime, R_t is a linear function of CBR_t, as in

$$R_t = \beta_1 + \beta_2 CBR_t + \varepsilon_t \tag{5}$$

where β_1 and β_2 are parameters, and ε_t is a disturbance term. Note that both R_t and CBR_t are measured as percentages, so their range of variation can be taken to be between 0 % and 100 %.

Formally, the mandate of the Norwegian Central Bank changed to inflation targeting on 29 March 2001, but the general understanding is that in practice inflation targeting was partly in operation long before. In order to test that belief we estimate (5) on three different samples of quarterly data. The first sample is from 1997, first quarter, to 2001, first quarter (i.e., 1997q1-2001q1). The second sample is 2001q2-2008q4. Finally we estimate (5)o n the full sample; 1997q1-2008q4. Below we have listed sample sizes (denoted T_1, T_2 and T), and the corresponding sum of squared OLS residuals, denoted RSS.

1997q1-2001q1:	$T_1 = 17$	$RSS_1 = 0.95$
2001q2-2008q4:	$T_2 = 31$	$RSS_s = 2.75$
1997q1-2008q4:	$T = 48$	$RSS = 3.85$

(a) Use this information to calculate a Chow-test statistic which is relevant for the hypothesis that the parameters of (5) are constant. What is your conclusion about the stability of (5)?

Under the H_0 we have $F(k, T-2k)$ where $k = 2$ due to the intercept and the derivative coefficient.

$$F(2, 44) = \frac{3.85 - (0.95 + 2.75)}{(0.95 + 2.75)} \times \frac{44}{2} = 0.89189$$

A colleague proposes to extend equation (5) with a variable ΔRW_t which is the change in the foreign money market interest rate, RW_t. (This interest rate is also measured in percentages). Her theory is that if the foreign interest rate is growing ($\Delta RW_t > 0$), the difference between R_t and CBR_t will increase, and that the difference will become smaller if ΔRW_t is negative. Another colleague points out that a relationship like (5) cannot be expected to hold during the financial crisis that hit the Norwegian money market in the fourth quarter of 2008. Based on this insight, an augmented model is estimated with the results reported in equation (6). $d2008q4_t$ is a dummy which is one in 2008q4, and zero otherwise. The numbers in round brackets below the parameter estimates are the corresponding standard errors.

$$
\begin{aligned}
R_t \;=\;& \underset{(0.0683)}{0.2155} + \underset{(0.0137)}{1.024}\, CBR_t + \underset{(0.082)}{0.4748}\, \Delta RW_t \\
(6) \qquad & + \underset{(0.185)}{1.606}\, d2008q4_t \\
& \text{OLS, } T = 48 \text{ (Sample is 1997q1-2008q4)} \\
& RSS = 1.31
\end{aligned}
$$

(b) Based on the evidence given in connection with question 2(a) above, and the results in equation (6), test the joint hypothesis that both the parameters of ΔRW_t and $d2008q_t$ are zero. You can base you answer on the assumption that the disturbances have the *classical properties* known from the textbooks.

$$F(2,44) = \left(\frac{3.85 - 1.31}{1.31}\right) \times \frac{46}{2} = 44.595$$

which can be compared to the critical values at the back of the set.

(c) Test also the hypothesis that a unit increase in CBR_t leads to a unit increase in R_t.
Using (6)

$$t(44) = \frac{(1.024 - 1)}{0.0137} = 1.7518.$$

which can be compared to the squarre root of F critical values at the back of the set.

5. The empirical relationship between the Central Bank's interest rate CBR_t and macroeconomic variables is also of interest. In an inflation targeting regime, the single most important variable is probably the annual inflation rate.We denote this variable by INF_t, and measure it as the 4 quarter percentage change in the consumer price index (adjusted for taxes and energy). The Norwegian Central Bank has communicated that (crisis situations aside) it will change the policy interest rate gradually and in small steps. This suggest that CBR_t can be modelled by an equation that includes CBR_{t-1} and INF_t as explanatory variables, and with $d2008q4_t$, to take account of the credit crises. Equation (7) gives the result from OLS estimation:

$$
\begin{aligned}
CBR_t \;=\;& \underset{(0.209)}{0.08755} + \underset{(0.0675)}{0.6939}\, CBR_{t-1} + \underset{(0.164)}{0.7752}\, INF_t \\
(7) \qquad & - \underset{(0.565)}{1.98}\, d2008q4_t \\
& \text{OLS, } T = 48 \text{ (Sample is 1997q1-2008q4)} \\
& R^2 = 0.93, \quad \hat{\sigma}_{OLS} = 0.54
\end{aligned}
$$

(a) Do you find the results in (7) statistically significant, and theoretically interpretable? (You can take as a given thing that the disturbances are homoscedastic and that there is no autocorrelation. R^2 is the multiple correlation coefficient. $\hat{\sigma}_{OLS}$ denotes the estimated residual standard error).

Significance of lagged CBR fits with gradualism in interest rate setting. That inflation is the intermediate target is confirmed by the significance of INF. If INF goes up 1 percentage point and stays at a higher level for some periods, CBR will increase by more than one percentage pointu, in accordance with the "Taylor principle" for example..

(b) One of your colleagues makes the claim that CBR_{t-1} is not an exogenous variable in (7). What would his argument be?

That CBR_{t-1} is correlated with past disturbances.

(c) A senior colleague says that in practice, you can regard CBR_{t-1} as exogenous, since the bias is probably of little numerical importance. However, she also says that a may be more important to look into the possibility that INF_t is an endogenous variable, but she expects that you can spell out the argument in more detail. What would your answer to this challenge be?

Simultaneity and/or that the true variable is INF_t^e so that INF_t contains an expectations error, equivalent to measurement errror.

(d) Together with your colleagues, you decide that U_t, the unemployment percentage, its lagged value, U_{t-1}, and lagged inflation INF_{t-1} are valid instruments for INF_t, and you therefore re-estimate the equation by 2SLS. The results are given in (8).

$$
\begin{aligned}
CBR_t &= \underset{(0.212)}{0.05501} + \underset{(0.195)}{0.9042}\ INF_t + \underset{(0.0757)}{0.652}\ CBR_{t-1} \\
&\quad - \underset{(0.575)}{2.079}\ d2008q4_t \qquad (8) \\
&\quad \text{2SLS, } T = 48 \text{ (Sample is 1997q1-2008q4)} \\
&\quad \hat{\sigma}_{2SLS} = 0.50\,,\ \mathsf{R}^2_{first-step} = 0.900
\end{aligned}
$$

Do you agree that this result is logically consistent with the policy communicated by the Norwegian Central Bank: namely that it primarily looks at the underlying inflation *tendency*, rather than at the actual rate of inflation, when the interest rate is decided? Explain. ($\hat{\sigma}_{2SLS}$ is the estimated residual standard error in (8), and $\mathsf{R}^2_{first-step}$ is the R^2 of the INF_t equation in the first step of the 2SLS procedure).

We interpret this model as having expected inflation INF_t^e, on the right hand side. The fitted values of INF_t from the first step regression are then the expected values of inflation. That the coefficient in (8) is larger than in (7) is consistent with the claim that the bank reacts to changes in predictable/underlying inflation.

(e) Finally, you want to investigate whether the central bank is "forward-looking" in its interest rate setting policy. Explain why one way of doing this is to replace INF_t by INF_{t+1}, and estimate with U_t, U_{t-1}, and INF_{t-1} as instruments for INF_{t+1}.

The results of this final equation is:

$$
\begin{aligned}
CBR_t &= -\underset{(0.215)}{0.1275} + \underset{(0.163)}{0.8686}\ INF_{t+1} + \underset{(0.0611)}{0.7007}\ CBR_{t-1} \\
&\quad - \underset{(0.552)}{2.03}\ d2008q4_t \qquad (9)\\
&\quad \text{2SLS, } T = 48 \text{ (Sample is 1997q1-2008q4)} \\
&\quad \hat{\sigma}_{2SLS} = 0.53,\ \mathsf{R}^2_{first-step} = 0.810
\end{aligned}
$$

On the basis of the above evidence, what is your conclusion about how forward-looking monetary policy has been?

The results are again consistent with the interpretation: the coefficient is almost unchanged, and is estimated with the same precision. If the Bank has a good forecast of a future shock to inflation, it will react to that challenge 3 months earlier than with (8). But if the shock is not precictable, it will not help.

Note to 5d) and 5e): Expect these to represent a challenge. There is no mathematically "correct" answer here. The only really disquailifying answer here is "have no comments". On the other hand, some of the students may write really good stuff?

(f) The CBR_t interest rate used above is the average interest rate in each quarter. Suppose instead that the dependent variable is *qualitative*, with 0 representing a decision (by the central bank) to lower the interest rate, 1 representing a decision to keep the policy interest rate constant, and 2 representing an increased central bank interest rate. Endogeneity issues aside, would you choose a linear regression model in this case?

In the same way as with a binary dependent variable, OLS is not statisfactory here since it will imply negative probabilities (or larger than 1) for some values of the explanatury variables. So would suggest LOGIT/PROBUIT, extended to the trinomial case.

Reference: 5% Critical values of $\mathbf{F}(v_1, v_2)$

		v_1	
v_2	1	2	...
: :	...	...	...
44	4.05	3.15	
:			