ECON 4160: ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

PROBLEM SET, EXAM AUTUMN 2010

Sensorveiledning/Assessment Guidance in italics

PROBLEM 1 (weight: 1/3)

In this problem we consider the two-equation model:

$$(1) \qquad y_{1t} = \beta_2 y_{2t} + \alpha_1 + u_{1t},$$
$$(2) \qquad y_{2t} = \beta_1 y_{1t} + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \alpha_2 + u_{2t},$$

where (y_{1t}, y_{2t}) are endogenous, (x_{1t}, x_{2t}) are exogenous and (u_{1t}, u_{2t}) are disturbances.

1A. Complete the model description, and show that equation (1) is identified. Explain briefly why satisfaction of the order condition for identification is equivalent to requiring that the number of instruments for the variables which need instruments should be sufficiently large.

Eq. (1) is overidentified. Order condition: We need at least one IV excluded from equation for each endogenous variable included in equation as RHS variable

1B. Explain briefly why the Ordinary Least Squares (OLS) estimator of β_2 obtained from (1) is inconsistent. How would you estimate β_2 consistently by using x_{1t} as an instrument for y_{2t}?

Compute plim of OLS estimator, and show that it is differs from β_2. $\widehat{\beta}_2^{IV} = M[y_1, x_1]/M[y_2, x_1]$.

1C. Two-stage least squares (2SLS) is another possible estimation method for β_2. Would you prefer it to the method proposed in **1B**? State the reason for your answer.

2SLS coincides with using $\widehat{y}_2$ as IV for y_2, and $\widehat{y}_2$ obtained from RF has higher correlation with y_2 than x_1 has. Hence, it gives lower asymptotic variance of estimator.

1D. Assume now that observations on x_{2t} are, for some reason, unavailable, although our theory claims that it should be included in equation (2). Assume also that x_{2t} is correlated with x_{1t}. In this situation it has been proposed to use as instrument for y_{2t} in equation (1), the variable $\widehat{y}_{2t}$ obtained as the fitted values in a regression of y_{2t} on x_{1t}. This is equivalent to use the fitted values from an OLS estimation of the reduced form equation for y_{2t} after having omitted x_{2t} from this reduced form equation.

Economists **a** and **b** discuss this method and state:

- **a:** "This method is quite OK and will produce a consistent estimator of β_2 even if x_{2t} is missing and omitted."
- **b:** "Since x_{2t} is omitted from the reduced form and is correlated with x_{1t}, the reduced form coefficient estimator will be inconsistent (omitted variables bias). This will bias the values of $\widehat{y}_{2t}$ obtained and therefore lead to inconsistent estimation of β_2 in equation (1)."

Do you agree with **a** or with **b**? State the reason for your answer.

***a** is correct: The IV used is a linear transformation of x_1 of the form $a + bx_1$. Biased estimation of RF equation for y_2 does NOT create biased IV estimator of SF coefficient. Formally: estimator invariant to (a, b). IV estimator invariant to one-to-one transformation of IV set.*

PROBLEM 2 (weight: 1/3)

We are concerned with the dynamic modelling of an aggregate consumption function, specified as a linear relation between the logarithm of private consumption, `y=ln(C)`, the logarithm of disposable income, `x=ln(R)`, and the real interest rate, `rint`. We have quarterly data from the U.S. and use observations from the period 1975.1–2000.4 (104 quarters) in the estimation. `C` and `R` are measured at constant prices, the real interest rate is measured as a decimal number (0.05 represents a 5 % pro anno rate, etc.). (The actual data series are longer than 104 quarters, up to three additional observations are used to construct lags.) Consider, for simplicity, `x` and `rint` as exogenous relative to `y`.

Edited printouts from PcGive for three dynamic specifications estimated by OLS are given below. In the second specification is used a smoothed log-income variable, `xs`, calculated from `x` as follows:

```
xs = x + 0.75*x(-1) + 0.50*x(-2) + 0.25*x(-3)
```

2A. Interpret the lag distributions involved in these three estimations.

EQ 1): Finite unrestricted lag distribution with current value and three lags for x and current value and one lag for rint. EQ 2): Finite restricted lag distribution across four periods (current and three lags) and with current value and one lag for rint, lag coefficients of x linearly declining. EQ 3): Infinite, geometric lag distribution of both x and rint with same λ parameter. Lag pattern described by three parameters only.

2B. Calculate from the printouts: (a) estimates of the short-run and the long-run elasticity of consumption with respect to income, (b) estimates of the short-run and the long-run effect of the interest rate on log-consumption.

Short-run: 1): 1.04 and -0.14, 2): 0.44 and -0.13, 3): 0.058 and -0.002.
Long-run: 1): 1.09 and -0.32, 2): 1.11 (=0.442 x 2.5) and -0.28, 3): 1.06(=0.0575/(1-0.948)) and -0.40 (=0.0020/(1-0.948)).

2C. The estimated short-run effects differ considerably across the specifications, also the t-values are markedly different. The estimated long-run effects are more similar. Can you give an interpretation of this finding? Supplementary statistics (for simplicity suppressed) indicate that the disturbances in all the three cases are autocorrelated.

Short-run: Strong multicollinearity between lagged values of x. Standard error estimates misleading because of autocorrelation. A somewhat more subtile point (but remember: this is no course in time-series econometrics!): Autocorrelated disturbances may 'disturb' the dynamic properties of the equations.

```
EQ(1) Modelling y by OLS. Estimation sample: 1975.1-2000.4 (104 observations)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant             -1.01331    0.05666    -17.9  0.0000   0.7673
x                     1.03887     0.1636     6.35  0.0000   0.2936
x_1                  0.210678     0.2090     1.01  0.3160   0.0104
x_2                -0.0593267     0.2094   -0.283  0.7775   0.0008
x_3                -0.0823840     0.1606   -0.513  0.6091   0.0027
rint                -0.139708    0.05732    -2.44  0.0166   0.0577
rint_1              -0.181896    0.05766    -3.15  0.0021   0.0930
------------------------------------------------------------------
sigma               0.0142624  RSS            0.0197314186
R^2                  0.996607  F(6,97) =      4749 [0.000]**
****************************************************************************

EQ(2) Modelling y by OLS. Estimation sample: 1975.1-2000.4 (104 observations)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant            -0.981466    0.06062    -16.2  0.0000   0.7238
xs                   0.442028   0.002894     153.  0.0000   0.9957
rint                -0.134583    0.06171    -2.18  0.0315   0.0454
rint_1              -0.148841    0.06155    -2.42  0.0174   0.0552
------------------------------------------------------------------
sigma               0.0153862  RSS             0.023673647
R^2                  0.995929  F(3,100) =     8156 [0.000]**
****************************************************************************

EQ(3) Modelling y by OLS. Estimation sample: 1975.1-2000.4 (104 observations)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant           -0.0454084    0.05228   -0.869  0.3871   0.0075
y_1                  0.948131    0.04590     20.7  0.0000   0.8101
x                   0.0575368    0.05080     1.13  0.2601   0.0127
rint              -0.00207814    0.02479  -0.0838  0.9334   0.0001
------------------------------------------------------------------
sigma              0.00645669  RSS           0.00416888716
R^2                  0.999283  F(3,100) =   4.647e+004 [0.000]**
```

PROBLEM 3 (weight: 1/3)

Consider the following three-equation model explaining two observable variables Y_1 and Y_2, for example the consumption of two commodities, by the same unobservable explanatory variable X^*, for example normal income, and an equation describing the relationship between the measured and the unobserved value of the explanatory variable:

$$(1) \qquad Y_1 = a_1 + b_1 X^* + v_1,$$
$$(2) \qquad Y_2 = a_2 + b_2 X^* + v_2,$$
$$(3) \qquad X = X^* + \varepsilon,$$

where X is the observed counterpart to X^* and ε is its measurement error. Assume that (v_1, v_2, ε) are (i) mutually uncorrelated, (ii) uncorrelated with X^* and (iii) have zero expectations and variances $(\sigma_1^2, \sigma_2^2, \sigma_\varepsilon^2)$, respectively. We have n observations of (Y_1, Y_2, X).

3A. Make the other assumptions you may find necessary and complete the model description. Would you say that (1)–(3) represent a system of regression equations? Explain briefly.

Strictly, a system of regression equations should have only observable variables. But the model could be argued to exemplify such a 3-equation system with X^* as a common latent exogenous variable.

3B. Derive from equations (1) and (3) a relationship between Y_1 and X. Next derive from this an expression for the asymptotic bias (inconsistency) of the OLS estimator of b_1.

OLS-estimator: $\tilde{b}_1 = M[Y_1, X]/M[X, X]$. *Ad plim: See textbook expositions.*

3C. Show that Y_2 satisfies the requirements for being an instrument for X in the equation obtained under **3B**, and write the expression for the estimator this motivates. Derive its probability limit to show that it is consistent.

Y_2 is correlated with X via X^*, and uncorrelated with $v_1 - b_1\varepsilon$.
IV-estimator: $\tilde{b}_1 = M[Y_1, Y_2]/M[X, Y_2]$

3D. In econometrics one often looks for 'good' instruments. Could you propose one or more criteria by means of which you can decide whether a proposed instrument, say Y_2 for X in question **3C**, is 'good' or 'bad'?

Correlation between instruments and instrumented variables. R^2 in reduced form equations. Tests for overidentifying restrictions.

3E. It has been argued that the disturbances v_1 and v_2 should be allowed to be correlated. If this is the case, would you then modify your conclusion under **3C**? Explain briefly.

This will 'destroy' the IV, as the mentioned correlation will make Y_2 correlated with $v_1 - b_1\varepsilon$.