ECON 4160: ECONOMETRICS –
MODELLING AND SYSTEMS ESTIMATION

PROBLEM SET, EXAM SPRING 2008

Sensorveiledning/Assessment Guidance in italics

PROBLEM 1 (weight: 60 %)

We are interested in analyzing, from micro data, the relationship between female labour supply, measured as the actual number of hours worked per year, and the length of education and work experience, measured in years. To explore this a cross-section data set from 753 females in the US observed in 1975 for the following six variables has been compiled:

```
Y1 = Number of hours worked in the year 1975
Y2 = Education, in years
X1 = Work experience, in years
Z1 = Father's education, in years
Z2 = Mother's education, in years
Z3 = Husband's education, in years
```

We assume that (`Y1`,`Y2`) are endogenous variables, that `X1` is exogenous, and that (`Z1`,`Z2`,`Z3`) have been proposed as candidates for being instruments for `Y2` in the equation

$$(*) \qquad \mathtt{Y1} = \alpha + \beta\,\mathtt{Y2} + \gamma\,\mathtt{X1} + \mathtt{U},$$

where `U` is a disturbance.

The estimation results and other printouts referred to below are obtained from PcGive and are given at the end of the problem.

(A): Give a stochastic specification of the model, and give reasons why treating (`Y1`,`Y2`) as jointly endogenous variables may be reasonable. In `EQ(1)`–`EQ(2)` two versions of $(*)$ are estimated, the first with γ set to zero *a priori*, the second with both coefficients free. Explain briefly why the two equations give different estimates of β and why both sets of OLS estimates are inconsistent.

Specify exogeneity. Other catchwords: Omitted regressor bias. Simultaneity bias.

(B): Assume that (`X1`,`Z1`,`Z2`,`Z3`) are exogenous variables in the model to which $(*)$ belongs and that the number of other equations and of other endogenous variables, say $N_*(\geq 1)$, is unknown. Show, by using the order condition, that $(*)$ is identified regardless of the value of N_*.

No. of excluded variables $= N_* + 3$. *This certainly exceeds No. of equations minus one* $= N_* + 2 - 1 = N_* + 1$ *for any* $N_*(\geq 1)$.

(C): Consider the estimates in printouts `EQ(3)` and `EQ(6)`. Explain briefly the terms `'IVE'` and `'Additional instruments'` and explain why the estimates are both consistent under the given assumptions. Why do they differ when computed from the 753 observations?

Full IV set =Included Exogenous variable(s) plus 'Additional instruments'. Different consistent estimates usually lead to different estimates when estimation sample is finite.

(D): Could `X1` alone have served as an instrument for `Y2` in $(*)$? State briefly the reason for your answer.

No. X1 serves as IV for itself. Using X1 also as IV for Y2 will lead to two identical normal equations, which will violate the rank condition for the full IV matrix vis-a-vis the equation's RHS variable matrix.

(E): Let (`Q1`, `Q2`, `Q3`, `Q4`) be four derived variables defined and calculated by PcGive by

```
Algebra code for variable transformations:
        Q1 = X1+Z1;
        Q2 = X1+2*Z1;
        Q3 = X1+Z2;
        Q4 = X1+2*Z2;
```

Explain why `(Q1,Q2,Q3,Q4)` are all valid instruments for $(*)$ and why the estimates in `EQ(4)`–`EQ(5)` coincide with those in `EQ(3)` and why the estimates in `EQ(7)`–`EQ(8)` coincide with those in `EQ(6)`.

[**Hint:** Note that (i) both `(X1,Q1)` and `(X1,Q2)` are one-to-one (non-singular) transformations of `(X1,Z1)` and (ii) `(X1,Q3)` and `(X1,Q4)` are one-to-one (non-singular) transformations of `(X1,Z2)`.]

Maybe this is a somewhat difficult (and unexpected) question, but it should be rather easy to prove by using matrix algebra: The equation is exactly identified, so IV=AZ [A quadratic and non-singular] will give the same IV estimator as IV=Z for any A. Also candidates unfamiliar with matrix algebra should have a change by noting that in all four cases X1 acts as IV for itself (perfect correlation) and "the rest of the IV set is disposed of as IV for Y2" Also candidates knowing that OLS is invariant to non-singular variable transformations while exploiting the relationship between IV and 2SLS could take advantage of this knowledge.

(F): Explain briefly the estimation method used for equations `EQ(9)`–`EQ(10)`, in particular how it differs from the methods used for equations `EQ(3)` and `EQ(6)`. Which conclusions do you draw from printouts `EQ(11)`–`EQ(12)` and the correlation matrix below about the quality of the instruments? What would you conclude about the effect on the female labour supply of (a) a one year increase in education, (b) a one year increase in work experience?

The catchwords here are overidentification and 2SLS as well as the R-square for the reduced for equation for Y2 as an overall IV quality index. The low t-value of Y2 and the high t-value of X1 should be noted.

(G): Good arguments may be given for treating `X1` as an endogenous variable, jointly determined with `Y1` and `Y2`. If you accept this, how would you then modify your model and proceed to estimate the coefficients of $(*)$? Explain briefly.

IVs will be needed for both Y2 and X1. The Zs are still candidates. But it may be remarked that in order to tackle this question, the model should probably be extended and more exogenous variables introduced.

PCGIVE PRINTOUTS FOR PROBLEM 1

```
MEANS, STANDARD DEVIATION AND CORRELATIONS. THE SAMPLE IS: 1 TO 753

Means
       Y1           Y2           X1           Z1           Z2           Z3
   740.58       12.287       10.631       8.8088       9.2510       12.491


Standard deviations (using T-1)
       Y1           Y2           X1           Z1           Z2           Z3
   871.31       2.2802       8.0691       3.5723       3.3675       3.0208


Correlation matrix:
             Y1           Y2           X1           Z1           Z2           Z3
Y1       1.0000      0.10596      0.40496     0.013671     0.057864   -0.0096504
Y2      0.10596       1.0000     0.066256      0.44246      0.43534      0.61195
X1      0.40496     0.066256       1.0000    -0.078802    -0.082179    -0.036301
Z1     0.013671      0.44246    -0.078802       1.0000      0.57307      0.36670
Z2     0.057864      0.43534    -0.082179      0.57307       1.0000      0.32447
Z3   -0.0096504      0.61195    -0.036301      0.36670      0.32447       1.0000

********************************************************************************
```

```
EQ( 1) Modelling Y1 by OLS-CS.  The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Y2                    40.4890      13.87     2.92   0.004   0.0112
Constant              243.094      173.3     1.40   0.161   0.0026

sigma                 866.986  RSS                564499772
R^2                 0.0112276  F(1,751) =    8.528 [0.004]**
log-likelihood       -6161.52  DW                     0.973
no. of observations       753  no. of parameters          2
mean(Y1)              740.576  var(Y1)               758180

****************************************************************

EQ( 2) Modelling Y1 by OLS-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Y2                    30.3699      12.74     2.38   0.017   0.0075
X1                    43.1593      3.599     12.0   0.000   0.1609
Constant             -91.3922      161.3   -0.567   0.571   0.0004

sigma                 794.728  RSS                473694833
R^2                  0.170281  F(2,750) =    76.96 [0.000]**
log-likelihood       -6095.49  DW                      1.18
no. of observations       753  no. of parameters          3
mean(Y1)              740.576  var(Y1)               758180

****************************************************************

EQ( 3) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2              Y     38.9067      28.31     1.37   0.170
X1                    42.9995      3.632     11.8   0.000
Constant             -194.584      345.5   -0.563   0.574

sigma                 794.966  RSS                473978536
Reduced form sigma     796.74
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Z1

****************************************************************

EQ( 4) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2              Y     38.9067      28.31     1.37   0.170
X1                    42.9995      3.632     11.8   0.000
Constant             -194.584      345.5   -0.563   0.574

sigma                 794.966  RSS                473978536
Reduced form sigma     796.74
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Q1

****************************************************************
```

```
EQ( 5) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2             Y      38.9067      28.31     1.37   0.170
X1                    42.9995      3.632     11.8   0.000
Constant             -194.584      345.5   -0.563   0.574

sigma                 794.966  RSS                473978536
Reduced form sigma     796.74
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Q2

*****************************************************************

EQ( 6) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2             Y      79.0119      29.01     2.72   0.007
X1                    42.2486      3.667     11.5   0.000
Constant             -679.367      354.0    -1.92   0.055

sigma                 802.418  RSS                482905547
Reduced form sigma     793.73
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Z2

*****************************************************************

EQ( 7) Modelling Y1 by IVE-CS.  The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2             Y      79.0119      29.01     2.72   0.007
X1                    42.2486      3.667     11.5   0.000
Constant             -679.367      354.0    -1.92   0.055

sigma                 802.418  RSS                482905547
Reduced form sigma     793.73
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Q3

*****************************************************************

EQ( 8) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2             Y      79.0119      29.01     2.72   0.007
X1                    42.2486      3.667     11.5   0.000
Constant             -679.367      354.0    -1.92   0.055

sigma                 802.418  RSS                482905547
Reduced form sigma     793.73
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         3
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Q4

*****************************************************************
```

```
EQ( 9) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2            Y       22.6733      18.74     1.21   0.227
X1                    43.3034      3.610     12.0   0.000
Constant              1.64213      231.5  0.00709   0.994

sigma                 794.922  RSS                473925434
Reduced form sigma     794.47
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         5
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Z1
[1] = Z2
[2] = Z3
*****************************************************************

EQ(10) Modelling Y1 by IVE-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob
Y2            Y       22.6733      18.74     1.21   0.227
X1                    43.3034      3.610     12.0   0.000
Constant              1.64213      231.5  0.00709   0.994

sigma                 794.922  RSS                473925434
Reduced form sigma     794.47
no. of observations       753  no. of parameters          3
no. endogenous variables    2  no. of instruments         5
mean(Y1)              740.576  var(Y1)               758180

Additional instruments:
[0] = Q1
[1] = Q3
[2] = Z3
*****************************************************************

EQ(11) Modelling Y1 by OLS-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
X1                    44.5060      3.605     12.3   0.000   0.1692
Z1                  -0.568442      10.18  -0.0558   0.956   0.0000
Z2                    26.3366      10.63     2.48   0.013   0.0081
Z3                   -7.74773      10.43   -0.743   0.458   0.0007
Constant              125.588      138.8    0.905   0.366   0.0011

sigma                 794.472  RSS                472127057
R^2                  0.173027  F(4,748) =    39.13 [0.000]**
log-likelihood       -6094.24  DW                      1.17
no. of observations       753  no. of parameters          5
mean(Y1)              740.576  var(Y1)               758180

*****************************************************************


EQ(12) Modelling Y2 by OLS-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
X1                  0.0318243   0.007590     4.19   0.000   0.0230
Z1                   0.101756    0.02144     4.75   0.000   0.0292
Z2                   0.130410    0.02238     5.83   0.000   0.0434
Z3                   0.373721    0.02195     17.0   0.000   0.2793
Constant              5.17748     0.2921     17.7   0.000   0.2957

sigma                  1.6726  RSS               2092.60733
R^2                  0.464812  F(4,748) =    162.4 [0.000]**
log-likelihood       -1453.28  DW                         2
no. of observations       753  no. of parameters          5
mean(Y2)              12.2869  var(Y2)              5.19262
```

PROBLEM 2 (weight: 40 %)

(A): Consider the simple macro model

$$C_t = \alpha + \beta Y_t + u_t, \tag{1}$$

$$Y_t = C_t + I_t + G_t. \tag{2}$$

where Y_t (= GNP) and C_t (= Total Private Consumption) are endogenous, I_t (= Total Gross Investment) and G_t (= Total Public Expenditure) are exogenous variables, and u_t is a disturbance. Complete the model description and explain which of its equations can be identified from time series on (Y_t, C_t, I_t, G_t). The marginal propensity to consume, β, can be estimated consistently by instrumental variables in four different ways, by using as instruments for Y_t, respectively, (i) only I_t, (ii) only G_t, (iii) $I_t + G_t$, or (iv) both I_t and G_t. Which of alternatives (i)–(iv) would you prefer if you believe in this simple model? State the reason for your answer.

Equation (1) is (exactly) identified. Identification problems related to (2) should not be discussed! The best answer to the final question is probably (iii), since I_t and G_t enter the model's reduced form only via their sum. However, (iv) can also be defended if one chooses to neglect the property that that the reduced form equations for I_t and G_t variables have the same coefficients, but if the candidate chooses so, this should be motivated.

(B): An extended version of the macro model is also of interest:

$$C_t = \alpha_1 + \beta_1 Y_t + u_t, \tag{3}$$

$$I_t = \alpha_2 + \beta_2 (Y_t - Y_{t-1}) + \gamma_2 G_t + v_t, \tag{4}$$

$$Y_t = C_t + I_t + G_t, \tag{5}$$

where (4), with $\beta_2 > 0$, $\gamma_2 > 0$, represents a hypothesis that gross investment responds partly to the increase in GNP and partly to certain components of Total Public Expenditure, and v_t is a disturbance. Complete the model description also in this case. Decide which of the model's equations can be identified from time series of (Y_t, C_t, I_t, G_t). How would you now estimate the consumption function?

[**Hint:** In interpreting (4) and specifying the model stochastically, you may consider it as having the form

$$I_t = \alpha_2 + \beta_2 Y_t + \beta_3 Z_t + \gamma_2 G_t + v_t$$

with the linear restriction $\beta_3 = -\beta_2$ imposed and with $Z_t = Y_{t-1}$ considered as predetermined (with properties which in this context can be treated as coinciding with those of an exogenous variable).]

Eqs. (3) and (4) are both identified, by the order condition. The restriction $\beta_3 = -\beta_2$ should then be counted as a linear restriction, so that (4) has two restrictions. For (3), 2SLS estimation with (G_t, Y_{t-1}) treated as IVs for Y_t will do.

(C): The reduced form equation for Y_t, obtained by inserting (3) and (4) into the national budget identity (5) and solving for Y_t is (derivation not required)

$$Y_t = a + bG_t + cY_{t-1} + \varepsilon_t, \tag{6}$$

where

$$a = \frac{\alpha_1+\alpha_2}{1-\beta_1-\beta_2}, \quad b = \frac{1+\gamma_2}{1-\beta_1-\beta_2}, \quad c = -\frac{\beta_2}{1-\beta_1-\beta_2}, \quad \varepsilon_t = \frac{u_t+v_t}{1-\beta_1-\beta_2}.$$

Would you consider (6) as describing a lag distribution, and if so which form does it have? Assume that consistent estimates of $(\beta_1, \beta_2, \gamma_2)$, satisfying $\beta_2 < \frac{1}{2}(1-\beta_1) \Longrightarrow |c| < 1$, are available (you are not required to propose an estimation procedure). Explain how you from this information would estimate b and c consistently and explain briefly how you from the estimates obtained, symbolized by $\widehat{\ }$, would proceed to compute the effect of a one unit increase in G in a particular year

(a) on Y in the current year,
(b) on Y in the next year, and
(c) on Y in the long run, i.e., the sum of the effects in the current and all future years.

[**Hint:** To illustrate your points you may well use numerical values, say
$(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\gamma}_2) = (0.65,\ 0.1,\ 0.05) \implies (\widehat{b},\ \widehat{c}) = (4.2,\ -0.4).$]

Geometric lag distribution with negative ratio and hence oscillating signs of the coefficients. Use Stutsky's theorem to prove consistency.

*Answer to **(a)**:* $b = \dfrac{1+\gamma_2}{1-\beta_1-\beta_2} = 4.2.$

*Answer to **(b)**:* $b\,c = \dfrac{(-\beta_2)(1+\gamma_2)}{(1-\beta_1-\beta_2)^2} = -1.68.$

*Answer to **(c)**:* $\displaystyle\sum_{i=0}^{\infty} b\,c^i = \frac{b}{1-c} = \frac{1+\gamma_2}{1-\beta_1-\beta_2}\sum_{i=0}^{\infty}\left(\frac{-\beta_2}{1-\beta_1-\beta_2}\right)^i = \frac{1+\gamma_2}{1-\beta_1} = 3.0.$