ECON 4160: ECONOMETRICS –
MODELLING AND SYSTEMS ESTIMATION

PROBLEM SET, EXAM SPRING 2007

Sensorveiledning in italics

PROBLEM 1 (weight: 40 %)

A relationship is often assumed to exist between finished goods inventories (in Norwegian: lager av ferdigvarer) and sales of manufactured commodities, but economists do not agree on how they are related. Ideally, for any such good, the difference between production and sales should equal the increase in the inventories. To explore this issue we have collected a data set which contains these two variables and a few others. The data set consists of seasonally adjusted quarterly data from the US, in billions of 2000 dollars, from 1982:1 to 2001:4 ($T = 80$ observations) for the following four variables and their one-quarter differences:

```
HRAW  = Raw material inventories for the manufacturing sector
HWIP  = Work in progress inventories for the manufacturing sector
HFIN  = Finished goods inventories for the manufacturing sector
SMAN  = Real manufacturing sales
DHRAW = One-quarter difference in HRAW
DHWIP = One-quarter difference in HWIP
DHFIN = One-quarter difference in HFIN
DSMAN = One-quarter difference in SMAN
```

Estimation results, from PcGive, is given at the end of the problem.

(1A):

(a) Consider first the two Ordinary Least Squares (OLS) regressions in EQ(1.1) and EQ(1.2). Could you explain why they have the same R^2?

$R[X,Y]^2 \equiv R[Y,X]^2 \equiv R^2$

(b) The estimate of the coefficient of `HFIN` in EQ(1.2) is not very far from the inverse of the estimate of the coefficient of `SMAN` in EQ(1.1). On the other hand, the gap between the corresponding coefficient estimates in the differenced equations EQ(1.3) and EQ(1.5) is much larger. Give a brief explanation of this.

$\widehat{\beta}_{OLS;Y|X}/\widehat{\beta}_{OLS;X|Y} \equiv \{M[Y,X]/[M[X,X]\}/\{M[Y,Y]/M[X,Y]\} \equiv R[Y,X]^2 \equiv R^2$
$\Longrightarrow$ *Small/large discrepancy* $\Longleftrightarrow$ R^2 *is large/small*

(1B):

It has been suggested, as a way of examining whether `SMAN` or `HFIN` should be treated as exogenous, to compute (i) the correlation coefficient between `SMAN` and the residuals from EQ(1.1) and (ii) the empirical correlation coefficient between `HFIN` and the residuals from EQ(1.2) to see how they relate to the zero correlation assumption between disturbances and regressors in a well-specified classical OLS regression equation with stochastic regressors. Comment briefly on this suggestion.

This is a nonsensical idea because OLS residuals are, by construction, orthogonal to the regressors. Both correlations are zero.

(1C):

By comparing EQ(1.1) with EQ(1.3) and EQ(1.4), we note that when transforming the equation between `HFIN` and `SMAN` from levels to differences, the Durbin-Watson statistic (DW) is substantially increased. Explain this, perform the Durbin-Watson tests and state your conclusion. An extract from a table for the Durbin-Watson critical values is given below.

DW 5 % Critical Values (dL,dU).
T = No. of obs.; K = No. of coef. (incl. intercept)

T	K	dL	dU	T	K	dL	dU
73	2	1.59243	1.64788	77	2	1.60361	1.65614
73	3	1.56446	1.67681	77	3	1.57710	1.68348
73	4	1.53599	1.70667	77	4	1.55015	1.71166
74	2	1.59530	1.65001	78	2	1.60626	1.65812
74	3	1.56772	1.67852	78	3	1.58010	1.68509
74	4	1.53966	1.70793	78	4	1.55351	1.71287
75	2	1.59813	1.65209	79	2	1.60887	1.66006
75	3	1.57091	1.68020	79	3	1.58304	1.68667
75	4	1.54323	1.70920	79	4	1.55679	1.71407
76	2	1.60090	1.65413	80	2	1.61143	1.66197
76	3	1.57404	1.68185	80	3	1.58592	1.68823
76	4	1.54673	1.71043	80	4	1.56001	1.71526

$u_t = \rho u_{t-1} + \epsilon_t \implies \Delta u_t = \epsilon_t - (1-\rho)u_{t-1}$
Hence, $\epsilon \sim \mathsf{IID}(0,\sigma^2)$ & $\rho \approx 1 \implies \Delta u_t \approx\sim \mathsf{IID}(0,\sigma^2)$.

(1D):

(a) There is reason to claim that neither `SMAN` nor `HFIN` is exogenous, but determined jointly with other variables in a multi-equation model. If this is true, what would you say about the properties of the estimates in EQ(1.1)–EQ(1.6)? Explain, with this in mind, how you would interpret the printouts in EQ(1.7) and EQ(1.8).

EQ(1.1)–EQ(1.6) all suffer from simultaneity bias. EQ(1.7) and EQ(1.8) give 2SLS estimates on original and renormalized form and are both consistent if the instruments used are valid.

(b) Can you from the printout in EQ(1.7)–EQ(1.10) draw conclusions about the quality of `DHRAW` and `DHWIP` as instruments for `DHFIN` and `DSMAN`.

The R^2s in the reduced form equations EQ(1.9)–EQ(1.10) are indicators of the quality of the IVs; confer first IV requirement.

(c) When OLS in EQ(1.5)–EQ(1.6) is replaced with IVE in EQ(1.7)–EQ(1.8), the estimated coefficients of `DHFIN` increase. The values of `sigma` and `RSS` also increase. Do you find this reasonable?

It can be argued that the OLS estimators have a negative asymptotic bias, which 2SLS eliminates. By construction, OLS minimizes `RSS`*, so it is not surprising that OLS comes out with the lowest reported values of this statistic.*

(d) Would you, when performing IVE estimation, recommend that the equation between `DSMAN` and `DHFIN` is specified with the latter as left-hand side variable and the former as right-hand side variable, rather than the opposite, as in EQ(1.8)? State briefly the reasons for your answers.

*According to the answer to **(b)**,* `DHRAW` *and* `DHWIP` *are of higher quality as IVs for* `DHFIN` *than they are as IVs for* `DSMAN`*. Hence the first normalization, that in EQ(1.8), seems preferable.*

PCGIVE PRINTOUTS FOR PROBLEM 1

```
EQ(1.1) Modelling HFIN by OLS. The estimation sample is: 1982(1) to 2001(4)

                 Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant             13.7841      3.493     3.95   0.000   0.1665
SMAN                0.428138    0.01247     34.3   0.000   0.9379

sigma                4.88443  RSS                1860.89814
R^2                 0.937929  F(1,78) =      1179 [0.000]**
log-likelihood      -239.387  DW                      0.171
no. of observations       80  no. of parameters           2
mean(HFIN)           132.215  var(HFIN)             374.755
****************************

EQ(1.2) Modelling SMAN by OLS. The estimation sample is: 1982(1) to 2001(4)

                 Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant            -13.0273      8.527    -1.53   0.131   0.0291
HFIN                 2.19072    0.06381     34.3   0.000   0.9379

sigma                11.0488  RSS                9521.93313
R^2                 0.937929  F(1,78) =      1179 [0.000]**
log-likelihood      -304.688  DW                      0.172
no. of observations       80  no. of parameters           2
mean(SMAN)           276.619  var(SMAN)             1917.56
****************************

EQ(1.3) Modelling DHFIN by OLS. The estimation sample is: 1982(2) to 2001(4)

                 Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant            0.615119     0.1797     3.42   0.001   0.1321
DSMAN              0.0534312    0.04514     1.18   0.240   0.0179

sigma                1.47532  RSS                167.595971
R^2                0.0178742  F(1,77) =       1.401 [0.240]
log-likelihood      -141.804  DW                       1.39
no. of observations       79  no. of parameters           2
mean(DHFIN)         0.696646  var(DHFIN)            2.16008
****************************

EQ(1.4) Modelling DHFIN by OLS. The estimation sample is: 1982(2) to 2001(4)

                 Coefficient  Std.Error  t-value  t-prob Part.R^2
DSMAN               0.112638    0.04446     2.53   0.013   0.0760

sigma                1.57341  RSS                193.097317
log-likelihood      -147.399  DW                       1.31
no. of observations       79  no. of parameters           1
mean(DHFIN)         0.696646  var(DHFIN)            2.16008
****************************
```

```
EQ(1.5) Modelling DSMAN by OLS. The estimation sample is: 1982(2) to 2001(4)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              1.29278     0.4596     2.81   0.006   0.0932
DHFIN                0.334528     0.2826     1.18   0.240   0.0179

sigma                 3.69152  RSS               1049.30278
R^2                 0.0178742  F(1,77) =       1.401 [0.240]
log-likelihood        -214.26  DW                      1.27
no. of observations        79  no. of parameters          2
mean(DSMAN)           1.52582  var(DSMAN)            13.524
*****************************

EQ(1.6) Modelling DSMAN by OLS. The estimation sample is: 1982(2) to 2001(4)

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
DHFIN                0.674971     0.2664     2.53   0.013   0.0760

sigma                 3.85159  RSS               1157.11119
log-likelihood       -218.123  DW                      1.26
no. of observations        79  no. of parameters          1
mean(DSMAN)           1.52582  var(DSMAN)            13.524
*****************************

EQ(1.7) Modelling DSMAN by IVE. The estimation sample is: 1982(2) to 2001(4)

                  Coefficient  Std.Error  t-value  t-prob
Constant             0.829679     0.5407     1.53   0.129
DHFIN          Y     0.999280     0.4707     2.12   0.037

sigma     3.82186   RSS     1124.71058   Reduced form sigma     3.5899
no. of observations        79  no. of parameters          2
no. endogenous variables    2  no. of instruments         3
mean(DSMAN)           1.52582  var(DSMAN)            13.524
Additional instruments:    [0] = DHRAW    [1] = DHWIP
*****************************

EQ(1.8) Modelling DSMAN by IVE. The estimation sample is: 1982(2) to 2001(4)

                  Coefficient  Std.Error  t-value  t-prob
DHFIN          Y      1.24113     0.4064     3.05   0.003

sigma     3.96151   RSS     1224.09962   Reduced form sigma     3.7617
no. of observations        79  no. of parameters          1
no. endogenous variables    2  no. of instruments         2
mean(DSMAN)           1.52582  var(DSMAN)            13.524
Additional instruments:   [0] = DHRAW     [1] = DHWIP
*****************************
```

EQ(1.9) Modelling DHFIN by OLS. The estimation sample is: 1982(2) to 2001(4)

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	0.366541	0.1414	2.59	0.011	0.0812
DHRAW	0.505126	0.09277	5.44	0.000	0.2806
DHWIP	0.196277	0.08548	2.30	0.024	0.0649

sigma	1.17377	RSS	104.707949
R^2	.386403	F(2,76) =	23.93 [0.000]**
log-likelihood	-123.224	DW	2.07
no. of observations	79	no. of parameters	3
mean(DHFIN)	0.696646	var(DHFIN)	2.16008

****

EQ(1.10) Modelling DSMAN by OLS. The estimation sample is: 1982(2) to 2001(4)

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	1.26427	0.4325	2.92	0.005	0.1011
DHRAW	0.270277	0.2837	0.953	0.344	0.0118
DHWIP	0.526503	0.2614	2.01	0.048	0.0507

sigma	3.58988	RSS	979.429648
R^2	0.083274	F(2,76) =	3.452 [0.037]*
log-likelihood	-211.538	DW	1.38
no. of observations	79	no. of parameters	3
mean(DSMAN)	1.52582	var(DSMAN)	13.524

END OF PRINTOUTS FOR PROBLEM 1

PROBLEM 2 (weight: 30 %)

Consider an econometric two-equation model with equations of the form:

$$(1) \qquad y = a + bx \qquad + u,$$
$$(2) \qquad x = c + dy + ez + v,$$

where (y, x, z) are variables, (a, b, c, d, e) are constants, (u, v) are disturbances with (unknown) variances $\sigma_{uu} > 0$, $\sigma_{vv} > 0$ and covariance σ_{uv}. We are in particular interested in estimating b consistently.

Specify which variables are exogenous and endogenous, and explain whether b is identified. If b is identified, how you would estimate it in the following six cases. Answer very briefly, in only one or two sentences in each case.

(2A): (y, x, z) are all observable;
(a, b, c, d, e) are unknown;
$cov(z, u) = \text{cov}(z, v) = 0$;
σ_{uv} is unknown.

(2B): (y, x, z) are all observable;
$e = 0$, (a, b, c, d) are unknown;
$cov(z, u) = \text{cov}(z, v) = 0$;
σ_{uv} is unknown.

(2C): (y, x, z) are all observable;
$d = 0$, (a, b, c, e) are unknown;
$cov(z, u) = \text{cov}(z, v) = 0$;
$\sigma_{uv} = 0$.

(2D): (y, z) are observable, x is not observable;
$d = 0$, (a, b, c, e) are unknown;
$cov(z, u) = \text{cov}(z, v) = 0$;
$\sigma_{uv} = 0$,

(2E): (y, z) are observable, x is not observable;
$c = d = 0, e = 1$, (a, b) are unknown;
$cov(z, u) = \text{cov}(z, v) = 0$;
$\sigma_{uv} = 0$

(2F): (y, z) are observable, x is not observable;
$c = d = 0, e = 1$, (a, b) are unknown;
$cov(x, u) = \text{cov}(x, v) = 0$;
$\sigma_{uv} = 0$

(2A): *b identified, by order condition. Use z as IV for x.*

(2B): *b is non-identified, by order condition.*

(2C): *The model is recursive. b can be consistently estimated by OLS on (1). Hence, b is identified.*

(2D): *Eliminating x we see that b is not identified.*

(2E): *Similar to a standard measurement error model, z being the observed counterpart to x and v being the measurement error. However, since cov$(z, u) = \text{cov}(z, v) = 0$ is assumed (NB: here we make a non-standard assumption for an EIV model), OLS on (1) with x replaced by z is consistent for b.*

(2F): *We here have strictly a standard measurement error (EIV) model, z being the observed counterpart to x and v being the measurement error. Since cov$(x, u) = \text{cov}(x, v) = 0$, OLS on (1) with x replaced by z is inconsistent for b. Equation (1) is non-identified.*

PROBLEM 3 (weight: 30 %)

We are interested in examining how females' decisions work or not depends on age, education, work experience and some other socioeconomic variables. The data set – from 1975 for $n = 753$ females in the US – contains the following 8 variables:

```
DUMW  = Dummy variable = 1 if female worked in 1975, else 0
AGE   = Female's age, in years
AGESQ = Female's age squared
EDU   = Female's educational attainment, in years
WEXP  = Female's previous labor market experience, in years
FAEDU = Father's educational attainment, in years
MOEDU = Mother's educational attainment, in years
CIT   = Dummy variable = 1 if female lives in a large city, else 0
```

The vector $\boldsymbol{x} = [\texttt{AGE}, \texttt{AGESQ}, \texttt{EDU}, \texttt{WEXP}, \texttt{FAEDU}, \texttt{MOEDU}, \texttt{CIT}]$, contains the variables to be treated as exogenous in the analysis below.

(3A):

Estimation results from OLS regression of DUMW on $\boldsymbol{x}$ is given in EQ(3.1) in the printout. Explain what you can conclude about the effects of a one year longer education period and a one year longer working experience on the females' propensity to work.

Equivalent to a linear probability model. Interpretation problematic because of the arbitrary metric of DUMW. Can identify the sign of the responses.

(3B):

(a) Logit and Probit models are used more frequently than linear regression models in analyzing individuals' discrete choice. Logit and Probit estimation results relating to female labour market responses are given in CS(3.2) and CS(3.3) in the printouts. Interpret these results, and in particular explain what you conclude about the effect on the females' propensity to work of

(i) a one year increase in the education period,
(ii) a one year increase in the working experience.

The coefficient vector in Logit model measures the effect of the covariates on the log-odds ratio. May, in both models, also be interpreted as the effect of the covariates on the underlying latent random utility.

(b) Could you explain why the Logit estimates are substantially higher (in absolute value) than the corresponding Probit estimates, even if the underlying problem is the same?

Difference variance of disturbance in the underlying utility equation. $\pi^2/3$ *in Logit,* 1 *in Probit. Rescaling convenient for comparison.*

(3C):

About 57 % of the females in the sample are working. Take $\bar{P}=0.5$ as a rough estimate of the probability of being employed. Can you – from the Logit results – estimate how the probability that a female will be working is affected by a one year increase in

(i) her education period,
(ii) her period of past working experience,
(iii) her father's education period and
(iv) her mother's education period?

Marginal effects at sample mean
= Vector of derivatives of response probabilities
$= P(1-P)\beta \approx \beta/4.$

PCGIVE PRINTOUTS FOR PROBLEM 3

```
EQ(3.1) Modelling DUMW by OLS-CS. The estimation sample is: 1 to 753

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant            -0.903595     0.4827    -1.87   0.062   0.0047
AGE                 0.0525896    0.02231     2.36   0.019   0.0074
AGESQ            -0.000745911  0.0002573    -2.90   0.004   0.0112
EDU                 0.0289180   0.008401     3.44   0.001   0.0157
WEXP                0.0247053   0.002176     11.4   0.000   0.1475
FAEDU              0.00448683   0.006211    0.722   0.470   0.0007
MOEDU             -0.00161627   0.005842   -0.277   0.782   0.0001
CIT                -0.0187213    0.03509   -0.534   0.594   0.0004

sigma                0.448639  RSS               149.951028
R^2                  0.188259  F(7,745) =    24.68 [0.000]**
log-likelihood       -460.881  DW                     0.351
no. of observations       753  no. of parameters          8
mean(DUMW)           0.568393  var(DUMW)           0.245322
*****************************
```

CS(3.2) Modelling DUMW by Logit. The estimation sample is: 1 to 753

	Coefficient	Std.Error	t-value	t-prob
Constant	-7.60072	2.497	-3.04	0.002
AGE	0.287136	0.1163	2.47	0.014
AGESQ	-0.00399504	0.001362	-2.93	0.003
EDU	0.148748	0.04359	3.41	0.001
WEXP	0.125311	0.01290	9.72	0.000
FAEDU	0.0208852	0.03126	0.668	0.504
MOEDU	-0.00911546	0.02897	-0.315	0.753
CIT	-0.0947049	0.1773	-0.534	0.593

log-likelihood	-436.246857	no. of states	2
no. of observations	753	no. of parameters	8
mean(DUMW)	0.568393	var(DUMW)	0.245322

BFGS estimation (eps1=0.0001; eps2=0.005): Strong convergence

	Count	Frequency	Probability	loglik
State 0	325	0.43161	0.43160	-229.5
State 1	428	0.56839	0.56840	-206.8
Total	753	1.00000	1.00000	-436.2

****

CS(3.3) Modelling DUMW by Probit. The estimation sample is: 1 to 753

	Coefficient	Std.Error	t-value	t-prob
Constant	-4.55365	1.485	-3.07	0.002
AGE	0.172773	0.06895	2.51	0.012
AGESQ	-0.00240777	0.0008037	-3.00	0.003
EDU	0.0888289	0.02588	3.43	0.001
WEXP	0.0740782	0.007302	10.1	0.000
FAEDU	0.0131296	0.01876	0.700	0.484
MOEDU	-0.00610996	0.01744	-0.350	0.726
CIT	-0.0440375	0.1061	-0.415	0.678

log-likelihood	-436.702481	no. of states	2
no. of observations	753	no. of parameters	8
mean(DUMW)	0.568393	var(DUMW)	0.245322

BFGS estimation (eps1=0.0001; eps2=0.005): Strong convergence

	Count	Frequency	Probability	loglik
State 0	325	0.43161	0.43167	-229.6
State 1	428	0.56839	0.56833	-207.1
Total	753	1.00000	1.00000	-436.7

END OF PRINTOUTS FOR PROBLEM 3