Mid-Term 2 2004 – Answers

1.
a) See lecture notes.Remember that $Var(b_{OLS}) = \sigma^2 (X'X)^{-1}X'\Omega X(X'X)^{-1}$
and efficient variance given by GLS: $Var(b_{GLS}) = \sigma^2 (X'\Omega X)^{-1}$ (***not*** $\sigma^2 (X'X)^{-1}$)

b) See lecture notes

c) Outline either Goldfeld-Quandt test or Breusch-Pagan or White

2. Given the following model, (in mean deviation form), you suspect that the variable, x_1, is measured with error.

$$y_i = bx_1 + u_i \qquad (1)$$

Outline the consequences for OLS estimation of the coefficient b in (1)
(7 marks)

Given a possible instrument x_2, let W = [y : x_1 : x_2] and

$$W'W = \begin{bmatrix} 66 & 8 & 3 \\ 8 & 4 & 1 \\ 3 & 1 & 3 \end{bmatrix}$$

the sample size is 100

a) see lecture notes. Main point is that measurement error causes attenuation bias (coefficients biased toward zero)

b) Find the OLS and IV estimates of the coefficient on x_1
(4 marks)

$b^{OLS} = \Sigma x_1 y / \Sigma x_1^2$ *and* $b^{IV} = \Sigma x_2 y / \Sigma x_1 x_2$

From matrix we know $W'W = \begin{bmatrix} y'y & y'x_1 & y'x_2 \\ x_1'y & x_1'x_1 & x_1'x_2 \\ x_2'y & x_2'x_1 & x_2'x_2 \end{bmatrix}$

So $b^{OLS} = \Sigma x_1 y / \Sigma x_1^2 = 8/4 = 2$

and $b^{IV} = \Sigma x_2 y / \Sigma x_1 x_2 = 3/1 = 3$

(so OLS estimate is attenuated toward zero as expected in presence of measurement error)

c) The estimated residual variance from the IV estimation and hence the standard error of the IV estimate of b

(8 marks)

Need $s^2_{IV} = u_{IV}' u_{IV} / n = (y - x_1 b_{IV})' (y - x_1 b_{IV}) / n = (y'y - 2y' x_1 b_{IV} + b_{IV}' x_1' x_1 b_{IV})/n$

From matrix and answer to above

$$s^2_{IV} = (66 - 2(8) + 3(4)3) / 100 = 86/100 = 0.86$$

So $Var(b_{IV}) = s^2_{IV} (Z'X)^{-1}(X'X)(Z'X)^{-1} = s^2_{IV} (\Sigma x_1 x_2)^{-1} \Sigma x_1^2 (\Sigma x_1 x_2)^{-1}$
$= 0.86(1)4(1) = 3.44$

So $SE(b_{IV}) = \sqrt{3.44} = 1.85$

(Hence IV estimate of b is not statistically significant at 5% level)

d) Outline a possible test for endogeneity of x_1 in (1)

(6 marks)

Wu-Hausman test – compare IV and OLS estimates accounting for sampling variation, Under null, IV & OLS consistent but OLS efficient. Under alternative only IV consistent. Reject null if estimated Wu-Hausman exceeds critical value An asymptotic equivalent version of the test is to take the residuals from the 1st stage of the regression (the regression of the endogenous variable on its instruments), and include them as additional regressor(s) in the original OLS equation (include an estimated residual for each endogenous rhs variable). Reject null (of no endogeneity) if t value(s) on added residuals is significant.

3. Determine the rank and order conditions for identification of each of the equations in the following model:

$$Y_1 + b_{12}Y_2 + g_{11}X_1 = u_1$$
$$b_{21}Y_1 + Y_2 + g_{21}X_1 + g_{22}X_2 + g_{23}X_3 = u_2$$

where Y_1 and Y_2 are the endogenous variables in the system and X_1, X_2 and X_3 are the exogenous variables

(20 marks)

Does this mean you can estimate either equation of the model using
a) OLS?
b) Instrumental variables? If so what instruments would you use?

(5 marks)

Order condition states that

$$K - k \geq g - 1$$

(number excluded exogenous ≥ number included endogenous less one)

Rank condition states that equation is identified if rank(AΦ) = G-1 and over-identified if there exists more than one sub-matrix of order G-1 in AΦ

Consider each equation in turn

In (1) K=3 k=1 g=2

So 3-1 > 2 -1

And equation is over-identified according to order condition

$$\text{Now } A\Phi = \begin{bmatrix} b_{11} & b_{12} & g_{11} & g_{12} & g_{13} \\ b_{21} & b_{22} & g_{21} & g_{22} & g_{23} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} g12 & g13 \\ g22 & g23 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ g22 & g23 \end{bmatrix}$$

So matrix has more than one sub-matrix of order G-1 = 1 (g_{22} and g_{23}) so equation is over-identified according to the rank condition

In (2) K=3 k=3 g=2

So 3-3 < 2 -1

And equation is not identified according to order condition

$$\text{Now } A\Phi = \begin{bmatrix} b_{11} & b_{12} & g_{11} & g_{12} & g_{13} \\ b_{21} & b_{22} & g_{21} & g_{22} & g_{23} \end{bmatrix} * 0 = 0$$

ie no restrictions so matrix does not exist. Equation is not identified according to the rank condition

b) Since (2) not identified no exogenous variable excluded from (2) to use as potential instrument for Y_1
However could use either X_2 or X_3 as instruments for Y_2 in (1) since correlated with Y_2 according to (2) but unrelated to Y_1 and u_1 according to (1).

*Will be more efficient to use **both** X_2 and X_3 as instruments **if** sample size is large enough. Otherwise small sample bias increases with number of instruments.*

4.

a) Key assumption that distinguishes fixed and random effects is whether observed covariates X are correlated with unobserveables (as in fixed effect which makes unobservables a vector of individual-specific dummy variables) or uncorrelated (as in random effects – which puts unobserveables into error term)

b) See Problem set 8 question 4

c) If T>=3, whether to use 1st diffs. Or within-groups depends on extent of autocorrelation in (original) residuals (both techniques are consistent with T fixed as N →∞).
If original residuals are uncorrelated, 1st differencing generates autocorrelation in the residuals and gives inefficient standard errors, so prefer within-groups.
If original residuals are correlated, 1st differencing can reduce extent of autocorrelation so might be an improvement on within-groups.
If T is large wrt N, could have non-stationary series in which case first differencing is more likely to avoid problem of spurious regression

***d)** Wu-Hausman test, this time comparing fixed effects (always consistent, but inefficient if random effects assumption valid) with random effects estimator (consistent only if random effects assumption of no correlation between unobserveables and X covariates is valid). If estimate exceeds critical value, reject null that r.e model is valid.*
Remember degrees of freedom for test is k-1 (ie one less than no. of rhs coefficients)