Tripos, Part 2A, Paper 3
Supervision 4

Question 1
Consider the following bivariate linear regression

$$y = \alpha + T\beta + u \tag{1}$$

where T is a binary treatment regressor. α and β are unknown parameters, and u is an error term

a. Describe in two sentences an empirical, real-life example where such an equation might arise.

b. Why might u be heteroskedastic in your example?

c. Why might T be endogenous in your example?

d. Suppose a single instrument z is available. Show that the population coefficient β satisfies

$$\beta = Cov(z, y)/Cov(z, T) \tag{2}$$

where $Cov(z, y)$ and $Cov(z, T)$, denote, respectively, the population covariance between z and y, and z and T. How can you use this information to obtain a consistent estimate of β?

e. Can you give an example of an instrument in your example? Argue why it might be a sensible IV.

Question 2
Consider the following wage equation that explicitly recognises that ability affects $\log(wage)$

$$\log(wage) = \alpha + \beta_1 educ + \beta_2 ability + u.$$

The above model shows explicitly that we would like to hold ability fixed when measuring the returns to education. Assuming that the primary interest is in obtaining a reliable estimate of the slope parameters β_1, and that there is no direct measurement for ability, explain how you would do this using a method based upon a proxy variable and an IV estimator. In doing so evaluate the following statement:

> *whilst IQ is a good candidate as a proxy variable for ability, it is not a good instrumental variable for educ.*

Question 3

Part B 2006 Exam Question

Consider the following linear model of log wages (w) explained using years of schooling (S), years of experience and its square (E, E^2), and 3 dummy variables indicating whether the individual was black (B), lived in the south (Sth), and lived in a metropolitan area (Sm).

$$w_i = \alpha + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 B_i + \beta_5 Sth_i + \beta_6 Sm_i + \varepsilon_i \tag{3}$$

ε_i is an unobserved error term, and i indexes individuals.

A reduced form model for schooling is written as

$$S_i = \delta + \mathbf{z}_i' \boldsymbol{\pi} + v_i, \tag{4}$$

where $\mathbf{z}_i$ is a $L \times 1$ vector including the instrument (lived near college) and all the exogenous variables in (3), $\boldsymbol{\pi}$ is a $L \times 1$ vector of unknown parameters, δ is an unknown scalar parameter, and v_i is an error term.

a) Why might $Cov(v_i, \varepsilon_i)$ be non-zero?

b) Assuming that S_i is endogeneous how might you use the reduced form equation in conjunction with (3) to identify the parameter β_1?

c) In Tables 1.1 and 1.2 we report results from estimating two models based on (3). Data is a sample of 3010 men taken from the US National Longitudinal Survey of Young Men; the year is 1976. Provide a careful comparison of the results based on the OLS and IV estimators. Discuss the choice of instruments.

Table 1.1 Wage Equation Estimated by OLS

Dependent: log(wage)

Variable	Estimate	Std. Error	t-ratio
constant	4.734	0.068	70.02
S	0.074	0.004	21.11
E	0.084	0.007	12.58
E^2	-0.002	0.000	-7.05
B	-0.189	0.018	-10.76
Sm	0.181	0.016	10.37
Sth	-0.125	0.015	-8.26

$R^2 = 0.291$, F=204.93

Table 1.2 Wage Equation Estimated by IV

Dependent Variable: log(wage)

Variable	Estimate	Std. Error	t-ratio
constant	4.066	0.609	6.68
S	0.133	0.051	2.58
E	0.056	0.026	2.15
E^2	-0.008	0.001	-0.06
B	-0.103	0.077	-1.33
Sm	0.108	0.005	2.17
Sth	-0.098	0.028	-3.41

Instrument: lived near college
used for: S
R^2 for reduced form for S: 0.119