Problem Set #3

ECONOMICS 240B
SPRING 2008

Part I: Due March 12
Part II: Due March 19

PART I: "Theoretical" questions:

Turn in (correct) answers to the following exercises from Ruud's text:

Chapter 20: Exercises 20.3, 20.7, 20.12
Chapter 24: Exercises 24.1, 24.3, 24.6.

Additional "Theoretical" Questions:

1. Suppose you had a two-equation linear model

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij}, \qquad i = 1, ..., N, \qquad j = 1, 2,$$

but you suspected the errors might be heteroskedastic for each equation, and the correlation in the errors across equations (but not across individuals) might vary across individuals, i.e.,

$$Cov(\varepsilon_{ij}, \varepsilon_{i'k}) = \begin{cases} 0 & \text{if } i \neq i'; \\ \sigma_{i,jk} & \text{if } i = i'. \end{cases}$$

(a) One special case of this setup would be a random coefficient model

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_{ij}, \qquad i = 1, ..., N, \qquad j = 1, 2,$$

with $\boldsymbol{\beta}_{ij}$ random across i and independent of the regressors $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$,

$$\begin{aligned} E[\boldsymbol{\beta}_{ij}] &\equiv \boldsymbol{\beta}_j, \\ C[\boldsymbol{\beta}_{ij}, \boldsymbol{\beta}_{i'k}] &= \begin{cases} \mathbf{0} & \text{if } i \neq i'; \\ \boldsymbol{\Gamma}_{jk} & \text{if } i = i', \end{cases} \end{aligned}$$

so that the original model holds with

$$\varepsilon_{ij} \equiv \mathbf{x}'_{ij}(\boldsymbol{\beta}_{ij} - \boldsymbol{\beta}_j).$$

Describe how you could construct Feasible GLS estimators of the regression coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for this model.

(b) If you were unwilling to impose the random coefficient structure on this two-equation system, but wanted to test an r-dimensional nonlinear hypothesis

$$H_0 : g(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbf{0}$$

on the regression coefficients for both equations, describe how you could construct an asymptotically-valid test for this hypothesis using the classical LS estimator

$$\hat{\boldsymbol{\beta}}_{LS} \equiv \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{y}_1 \\ (\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}_2\mathbf{y}_2 \end{pmatrix}.$$

Give algebraic expressions for the components of your test statistic and precisely state the critical region for the test.

2. Let $\mathbf{y}$ be an $(N \times 1)$ vector of dependent variables, $\mathbf{X}$ an $(N \times K)$ matrix of (possibly endogenous) regressors, and $\mathbf{Z}$ an $(N \times L)$ matrix of instrumental variables (with $L \geq K$). Define

$$\hat{\mathbf{X}} \equiv \mathbf{Z}\hat{\mathbf{\Pi}},$$

where $\hat{\mathbf{\Pi}}$ is the matrix of regression coefficients for the regression of $\mathbf{X}$ on $\mathbf{Z}$:

$$\hat{\mathbf{\Pi}} \equiv (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}.$$

Show that the following four definitions of the two-stage least squares estimator $\hat{\beta}_{2SLS}$ are algebraically identical:

(i) the instrumental variables coefficient estimator for $\mathbf{y}$ on $\mathbf{X}$ using $\hat{\mathbf{X}}$ as a matrix of instrumental variables;

(ii) the classical LS regression coefficients in the regression of $\mathbf{y}$ on $\hat{\mathbf{X}}$;

(iii) the classical LS regression coefficients in the regression of $\hat{\mathbf{y}}$ on $\hat{\mathbf{X}}$, where

$$\begin{aligned} \hat{\mathbf{y}} &\equiv \mathbf{Z}\hat{\boldsymbol{\pi}}, \\ \hat{\boldsymbol{\pi}} &\equiv (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}; \end{aligned}$$

(iv) the coefficients on $\mathbf{X}$ in the classical LS regression of $\mathbf{y}$ on $\mathbf{X}$ and $\hat{\mathbf{V}}$, where $\hat{\mathbf{V}}$ is the matrix of first-stage residuals

$$\hat{\mathbf{V}} \equiv \mathbf{X} - \hat{\mathbf{X}};$$

(v) the coefficients on $\mathbf{X}$ in the classical LS regression of $\hat{\mathbf{y}}$ on $\mathbf{X}$ and $\hat{\mathbf{V}}$.

PART II: "Empirical" question:

1. In the file "earnings.txt" are data taken from Table A.3 in the text *A Course in Econometrics* by A. Goldberger. This data set has $n = 100$ observations (each a row of the table); there are 12 variables in the columns of the table, which represent:

V1= ID number	V2 = Family size	V3 = Education	V4 = Age
V5 = Experience	V6 = Months worked	V7 = Race	V8 = Region
V9 = Earnings	V10 = Income	V11 = Wealth	V12 = Savings

Consider a linear model for months worked (V6) as a function of a constant term, monthly earnings (V9 divided by V6), and family size (V2), and compare the least-squares estimates of the coefficients of this model to the corresponding two-stage least squares estimates, using a constant term, family size (V2), education (V3), age (V4), and race (V7) as instrumental variables. Discuss the plausibilty of the exclusion restrictions and interpret the difference in the two sets of estimates (and their statistical significance) from an economic perspective.