Problem Set #3

ECONOMICS 240B
SPRING 2006

Due March 8

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.

Extra Question:. 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.