BOSTON COLLEGE
Department of Economics
EC 771: Econometrics
Spring 2009
Prof. Baum, Ms. Skira

Problem Set 2: due Tuesday 17 February 2009 at classtime

1. Given the three vectors x_1, x_2, x_3:

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Show that any vector $z = b_1 x_1 + b_2 x_2$ in the subspace $\delta(x_1, x_2)$ also belongs to the subspace $\delta(x_1, x_3)$ and $\delta(x_2, x_3)$, where the subspace $\delta(\cdot)$ of $\mathbb{R}^3$ is the vector space spanned by the arguments. Give explicit formulas for z as a linear combination of x_1, x_3 and of x_2, x_3.

2. Consider the two regressions

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$
$$y = \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + u$$

where $z_1 = x_1 - 2x_2$, $z_2 = x_2 + 4x_3$ and $z_3 = 2x_1 - 3x_2 + 5x_3$. Let $\mathbf{X} = [x_1\ x_2\ x_3]$ and $\mathbf{Z} = [z_1\ z_2\ z_3]$. Show that the columns of $\mathbf{Z}$ can be expressed as linear combinations of the columns of $\mathbf{X}$, that is, $\mathbf{Z} = \mathbf{XA}$ for some 3×3 matrix $\mathbf{A}$. Find the elements of this matrix.

Show that the matrix $\mathbf{A}$ is invertible by showing that the columns of $\mathbf{X}$ are linear combinations of the columns of $\mathbf{Z}$. Find the elements of $\mathbf{A^{-1}}$. Show that the two regressions give the same fitted values and residuals.

Precisely how is the OLS estimate b_1 related to the OLS estimate a_1?

3. Use the dataset `http://fmwww.bc.edu/ec-p/data/greene2008/tbrate` in Stata for this exercise. It contains data (1950q1-1996q4) on three series: r_t, the three-month Treasury bill rate, π_t, the rate of inflation and y_t, the log of real GDP. For the period 1950q4–1996q4, run the regression

$$\Delta r_t = \beta_1 + \beta_2 \pi_{t-1} + \beta_3 \Delta y_{t-1} + \beta_4 \Delta r_{t-1} + \beta_5 \Delta r_{t-2} + u_t$$

(Hint: use the Stata timeseries operators `D.` and `L.`). Plot the residuals and fitted values against time (hint: `predict` and `tsline`). Regress the residuals on the fitted values and a constant. What do you learn from this second regression? Now regress the fitted values on the residuals and a constant. What do you learn from this third regression?

4. For the same sample period, regress Δr_t on a constant, Δy_{t-1}, Δr_{t-1} and Δr_{t-2}. Save the residuals from this regression (call them $\hat{e}_t$). Then regress π_{t-1} on the same regressors. Save the residuals from this regression and call them $\hat{v}_t$. Now regress $\hat{e}_t$ on $\hat{v}_t$. How are the estimated coefficients and residuals from this last regression related to anything that you obtained when you ran the regression in exercise 3?

5. For any pair of random variables b_1, b_2, show, by using the fact that the covariance matrix of $\mathbf{b} = (b_1\ b_2)$ is positive semidefinite, that

$$(Cov(b_1, b_2))^2 \leq Var(b_1)Var(b_2)$$

and show that the correlation between b_1, b_2 must lie between $-1, +1$.

5. Consider the linear regression model

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$$

Explain how you could estimate this model subject to the restriction that $b_2 + b_3 = 1$ by running a regression that imposes the restriction. Also explain how you could estimate the unrestricted model in such a way that the value of one of your coefficients would be zero if the restriction held exactly in your data.

6. Using Stata routine `ivreg2` (`ssc install ivreg2` if needed; full details on `ivreg2` in BC WP 667, (Baum, 2006) and `help ivreg2`):

```
use http://fmwww.bc.edu/ec-p/data/hayashi/griliches76.dta
xi i.year
```

a. Estimate the regression of log wage (lw) on experience (expr), years of schooling (s) and iq, considering iq as potentially mismeasured; instrument the equation with age, kww and med (mother's years of education). What is the identification status of this equation?

b. These data are pooled cross-section time-series (but not a panel). Introduce time effects (the year dummies) and reestimate the equation. What effect has this had on the model? What do you conclude?

c. Reestimate the equation of part b using robust standard errors. What effect does this have on the estimated model?

d. Reestimate the equation of part b using generalized method of moments (IV-GMM) with robust standard errors. What effect does this have on the model? What assumptions have been relaxed vis-a-vis the model estimated in part b? How do you interpret the Hansen J statistic for this model? How do you interpret the Anderson canonical correlation statistic?

e. Use the IV-GMM `endog` option (the "C" or GMM distance test) to test a subset of the orthogonality conditions: the exogeneity/endogeneity of s (years of schooling). What do you conclude?

f. Reestimate the model treating s as endogenous. What does the Hansen J test signify in this context?