Problem Set 2

EE426 Econometrics 2

Due February 24, 2014 (5pm)

Please report the regression results in each problem and print STATA .do file attached at the end of your answer.

1. In the model with one endogenous explanatory variable, one exogenous explanatory variable, and one extra exogenous variable, take the reduced form for y_2: $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$, and plug it into the structural equation: $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$. This gives the reduced form for y_1 : $y_1 = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + v_1$

(1.1) Find the α_j in terms of the β_j and the π_j.

(1.2) Find the reduced form error, v_1, in terms of u_1, v_2, and the parameters.

(1.3) How would you consistently estimate the α_j ?

2. The data in FERTILS2.dta for women in Botswana during 1988 includes information on number of children, years of education, age, and religious and economic status variables.

(2.1) Estimate the model

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

by OLS, and interpret the estimates (educ = mother's education, age = mother's age, children = number of children that mother has). Holding age fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

(2.2) The variable *frsthalf* is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that *frsthalf* is uncorrelated with the error term from (2.1), show that *frsthalf* is a reasonable IV candidate for *educ*. (Need to do a regression.)

(2.3) Estimate the model from (2.1) by using *frsthalf* as an IV for *educ*. Compare the estimated effect of education with the OLS estimate from (2.1)

(2.4) Which model should we rely on, OLS or IV? Show an appropriate test on how you select the model.

3. Use the data in CARD.dta. Card(1995) used a dummy variable for whether someone grew up near a four-year college (*nearc4*) as an instrumental variable for education. In a log(wage) equation, he included *educ*, *exper*, $exper^2$, a black dummy variable (*black*), dummy variabls for living in an SMSA and living in the South (*smsa* and *south*), and a full set of regional dummy variables and SMSA dummy for where the man was living in 1966 (*smsa66, reg662, ..., reg669*).

(3.1) Run the first stage regression (reduced form for *educ*), that is regress educ on *nearc4* and all of the exogenous variables (as mentioned above). Write the equation and report the results. Test whether *nearc4* can be used as an IV for *educ*.

(3.2) Run the OLS and IV estimations and report your results. Compare the estimates of the return to education between the two models.

(3.3) Obtain the reduced form residuals, $\hat{v}_2$, from (3.1). Use these to test whether educ is exogenous; that is determine if the difference between OLS and IV (as we see from 3.2) is statistically significant. (Show each step clearly on your answer.)

(3.4) Estimate the equation by 2SLS, adding *nearc2* as an instrument. Does the coefficient on *educ* change much?

(3.5) Test the overidentifying restriction from (3.4) (use Chi-Square distribution table at the end of your textbook, Table G4). Show each step clearly how do you test overidentifying restriction.