BOSTON COLLEGE
Department of Economics
EC771: Econometrics
Spring 2012
Prof. Baum, Mr. Simeonov

Problem Set 3: due Wednesday 14 March 2012 at classtime

1. Consider a simple consumption function of the form

$$c_i = \beta_1 + \beta_2 y_i^* + u_i^*,\ u_i^* \sim IID(0, \sigma^2)$$

where c_i is the log of consumption by household i and y_i^* is the unobservable permanent income of household i. Instead, we observe current income $y_i = y_i^* + \nu_i$, where $\nu_i \sim IID(0, \omega^2)$ is assumed to be uncorrelated with both y_i^* and u_i. We run the regression

$$c_i = \beta_1 + \beta_2 y_i + u_i$$

Under the plausible assumption that the population value of β_2 is positive, show that y_i is negatively correlated with u_i.

2. For this exercise `use` the data in

`http://fmwww.bc.edu/ec-p/data/dmackinnon/money`

Use `describe` and `tsset` for information about the dataset. Using these data, estimate the model

$$m_t = \beta_1 + \beta_2 r_t + \beta_3 y_t + \beta_4 m_{t-1} + \beta_5 m_{t-2} + u_t$$

by IV (`ivreg2`) for the period 1968q1–1998q4, treating r_t as endogenous, with two lagged values of r_t as excluded instruments. Perform a Durbin–Wu–Hausman test for the hypothesis that the equation must be estimated by instrumental variables rather than ordinary least squares. Perform a Sargan test for the overidentifying restrictions. Discuss your findings from these tests.

3. For this exercise `use` the data in

`http://fmwww.bc.edu/ec-p/data/dmackinnon/demand-supply`

The demand equation to be estimated is

$$q_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \gamma p_t + u_t$$

where q_t is the log of quantity, p_t is the log of price, x_{t2} is the log of income and x_{t3} is a dummy variable that accounts for regular demand shifts. Estimate this equation (a) with OLS and (b) with IV (`ivreg2`) using the variables x_{t4} and x_{t5} as instruments. Does OLS estimation appear to be valid? Does IV estimation appear to be valid? Perform appropriate tests to answer those questions.

Reverse the roles of q_t and p_t in the equation and estimate with OLS and IV. How are the two estimates of the coefficient of q_t in the new equation related to the corresponding estimates of γ in the original equation? What do these results suggest about the validity of the OLS and IV estimates?

4. With the Stata data set

`use http://fmwww.bc.edu/ec-p/data/wooldridge/wage2`

a. Estimate the regression of log wage on education, experience, tenure, and dummies for married, black, south, and urban. Ceteris paribus, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant?

b. Extend the model to allow the return on education to depend on race, and test that hypothesis.

c. Starting with the original model, allow wages to differ across four groups: married/black, married/nonblack, nonmarried/black and nonmarried/nonblack, taking account of the interactions between those characteristics. (Hint: factor variables make this much easier). What is the estimated wage differential between married blacks and married nonblacks? (Hint: `lincom` is useful in reporting these results).

5. With the Stata data set

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/gpa2
```

a. Consider the regression of *colgpa* on *hsize*, $hsize^2$, *hsperc*, *sat*, *female* and *athlete* where these variables are, respectively, college grade point average, size of high school graduating class, academic percentile in high school class, SAT score, and dummies for female and athlete. What are your expectations about the coefficients in this equation?

b. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

c. Drop *sat* from the model and reestimate the equation. Does this change the estimated effect of being an athlete?

d. In the original model, allow the effect of being an athlete to differ by gender, and test the hypothesis that there is no difference in estimated GPA between female athletes and female nonathletes.

e. Does the effect of *sat* significantly differ by gender?

6. With the Stata data set

```
use http://fmwww.bc.edu/ec-p/data/wooldridge2k/401ksubs
```

a. Test the hypothesis that average $nettfa$ does not differ by 401k eligibility status. What is the dollar amount of the estimated difference?

b. Estimate the regression of $nettfa$ on *inc*, inc^2, *age*, age^2, *male* and $e401k$. Are the quadratic terms justified? What is the estimated dollar effect of 401k eligibility?

c. Add the interaction terms $e401k(age - 41)$ and $e401k(age - 41)^2$ to the model. Given that the average age is about 41, the effect of $e401k$ is the estimated effect at the average age. Are these interaction terms significant? Interpret how they alter the model. Do the estimated effects of 401k eligibility at age 41 differ much between this model and the model of part b?

d. Drop the interaction terms from the model, but define five family size dummies, $fsize1$–$fsize5$ where $fsize5 = 1$ if $fsize \geq 5$. (This is most easily done by creating a new family size variable and then using factor-variables notation). Include these dummies in the model (choosing a base group). Are they jointly significant? How do you interpret the resulting regression?

e. Consider the regression of part b in its fully interacted ("Chow test") form, in which both intercepts and slopes are allowed to differ over the five family size categories. (Hint: you will definitely want to use factor variables to do this). Test the hypothesis that the regression function is stable over family size categories.