BOSTON COLLEGE
Department of Economics
EC771: Econometrics
Spring 2008
Prof. Baum, Mr. Alva

PROBLEM SET 5: DUE TUESDAY 29 APRIL 2008 AT CLASSTIME

1. Using the "canned" dataset `grunfeld`, which you may access from within Stata with the command

```
.webuse grunfeld
```

a. Run the regression

$$Invest_{it} = \beta_1 + \beta_2 Mvalue_{it} + \beta_3 Kstock_{it} + u_t$$

using firm fixed effects and random effects. Perform a Hausman test to determine which estimator is appropriate for this model.

b. Estimate the model via two-way fixed effects.

c. Estimate the original model employing the Seemingly Unrelated Regression estimator (`help sureg`). You will need to reshape the data into "wide" format. Discuss how this model differs from the fixed effects model estimated in part a. Test for cross–equation equality of the slope coefficients.

2. Use the "canned" dataset `pntsprd`, which you may access from within Stata with the command

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

You may want to

```
type http://fmwww.bc.edu/ec-p/data/wooldridge/pntsprd.des
```

a. The variable `favwin` is a binary variable which equals 1 if the team favored by the Las Vegas point spread wins the game. Estimate a linear probability model of `favwin` on `spread`. Explain why, if the spread incorporates all relevant information, we expect an intercept of 0.5. Test this hypothesis using robust standard errors. Is `spread` statistically significant? What is the estimated probability that the favored team wins when `spread`=10?

b. Estimate a probit model for Pr(favwin=1 — spread). Interpret and test the null hypothesis that the intercept is zero. What is the estimated probability that the favored team wins when `spread`=10? What is $\partial Pr(\cdot)/\partial spread$ when `spread`=10?

c. Add the variables `favhome, fav25, und25` to the probit model and test their joint significance. Does the spread incorporate all relevant information?

3. Use the "canned" dataset `loanapp`, which you may access from within Stata with the command

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

You may want to
`type http://fmwww.bc.edu/ec-p/data/wooldridge/loanapp.des`

a. Estimate a probit model of `approve` on `white`. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability model estimates?

b. Add the variables `hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, vr` to the probit model. Is there statistically significant discrimination against nonwhites?

c. Estimate this expanded model by logit and compare the coefficient on `white` to the probit estimate. Use the `mfx` command to calculate the effects of `white` on the probability of approval.

d. Use Tamas Bartus' `margeff` (`ssc install margeff`) on the probit form of the model. Why do the effects differ from those of `mfx`?

4. Use the "canned" dataset `fringe`, which you may access from within Stata with the command

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

You may want to
`type http://fmwww.bc.edu/ec-p/data/wooldridge/fringe.des`

a. For what percentage of workers in the sample is `pension` equal to zero? What is the range of `pension` for workers with nonzero pension benefits? Why is a Tobit model appropriate for modeling `pension`?

b. Estimate a Tobit model explaining `pension` in terms of `exper, age, tenure, educ, depends, married, white, male`. Do whites and males have statistically significant higher expected pension benefits?

c. Use the results from (b) to estimate the difference in expected pension benefits for a white male and a nonwhite female, both of whom are 35 years old, single with no dependents, with 16 years of education and 10 years of work experience. Hint: see `mfx`.

d. Add `union` to the Tobit model and comment on its significance.

5. Use the "canned" dataset `mroz`, which you may access from within Stata with the command

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

You may want to
`type http://fmwww.bc.edu/ec-p/data/wooldridge/mroz.des`

a. Using the 428 women who were in the workforce, estimate the return to education (`educ`) by OLS on `lwage` including `exper, expersq` as additional explanatory variables.

b. Estimate the return to education via two-step Heckit (hint: `help heckman`) where all exogenous variables show up in the second-stage regression, as well as `nwifeinc, age, kidslt6, kidsge6`. Is there evidence of selection bias? What parameter provides you with an indication of selection bias?

c. Reestimate the model with Heckman's maximum likelihood approach. Does any of your inference change? What parameter provides you with an indication of selection bias?

6. Use the "canned" dataset `cps91`, which you may access from within Stata with the command

```
use http://fmwww.bc.edu/ec-p/data/wooldridge2k/cps91
```

a. What fraction of working women report being in the labor force? Using only those data, estimate a wage equation explaining `lwage` as a function of `educ, exper, expersq, black, hispanic` with OLS. Do there appear to be significant wage differences by race/ethnicity?

b. Estimate a probit model for `inlf` that includes these explanatory variables as well as `nwifeinc, kidslt6`. Do these last two variables have the expected signs? Are they statistically significant? For the purposes of testing (and possibly correcting) for selection bias, why is it important that these two variables help explain `inlf`? What must you assume about these two variables in the `lwage` equation?

c. Compute the inverse Mills ratio for each observation and add it as an additional regressor in the wage equation from (a). What is its two-sided p-value? Does adding the IMR change the coefficients in the wage equation in important ways?