ECON 9730: Advanced Econometrics

Exercise #6
Due Tuesday, November 29 at the beginning of class[1]

Shif Gurmu

1. The text data set in **pension.raw** contains information on participant-directed pension plans for U.S. workers. These are part of the data originally used by Papke (2003) to estimate the effect of participant investment choice on asset allocation, among other things.[2] Variables are defined in the text file **pension.des**; see Papke's paper for further details. The key response variable of interest here is *pctstk*, which is asset allocation in defined contributions in pension plans and coded as 0, 50, 100 for responses "mostly bonds", "mixed", and "mostly stocks", respectively. As you can see from her paper, these categories were based on a latent variable representing percentage invested in stocks. The complete list of explanatory variables are *choice*, *age*, *educ*, *female*, *black*, *married, finc25*, ..., *finc101*, *wealth89*, and *prftshr*. The variable of most interest is *choice*, a dummy variable equal to 1 if the worker has a choice in how to allocate pension funds among different investments. The data and variable description files as well as the paper are zipped in file **papke.zip**.

A. Estimate an ordered probit model for *pctstk*, where the explanatory variables are as identified above.

B. The sample contains separate observations for some husband-wife pairs. Obtain the standard errors of the coefficient estimates from part (A) that are robust to cluster correlation within family. Do the robust standard errors differ much from the usual ordered probit standard errors, or from the heteroskedasticity-robust standard errors?

C. Obtain the estimated average marginal effects of *choice* on probabilities of asset allocation in defined contribution plans. Interpret the results.

D. Estimate the marginal effects of *choice* on probabilities of asset allocation in defined contribution plans for a single, nonblack female with 12 years of education who is 60 years old. Assume she has net worth (in 1989) equal to \$150,000 and earns \$45,000 per year, and her plan is not profit sharing.

E. Using the results and the person of part (D) above, estimate $E(pctstk|\mathbf{x})$. What can you say about the effect of choice for this person's investment in stocks?

Remarks - To get you started, please see **Example 16.2** on page 657 in Wooldridge (2010).

2. In this problem you will estimate the random effects probit as well as the dynamic random effects probit models of union membership using the Vella and Verbeek (1998)[3] data set. First, read the text file **readme-vv.txt** for description of the data. After unzipping the **vv-data.zip**, you also get the text data file **vv-data.dat**. Provide descriptive statistics for all variables in the data set.

A. Using the explanatory variables (including *occupational dummies* but excluding *lagged union status*) employed by Vella and Verbeek (1998) in modeling union membership (see Table II), obtain the **pooled probit** estimates. Find the average marginal effects of the following: *logexper*, *mar*, *exper*.

[1] This assignment can be done in groups of up to three students. You need to submit one solution per group.

[2] Papke, L.E. (2003), "Individual Financial Decisions in Retirement Saving Plans: the Role of Participant-direction," *Journal of Public Economics* 88, 39-61.

[3] Vella, F. and M. Verbeek (1998) "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men," *Journal of Applied Econometrics*, 13,163-183.

B. Restricting your sample period to 1981 - 1987, estimate the union membership model using the **random effects probit** estimator. Again, you should include occupational dummies but exclude *lagged union status.* Find the average marginal effect of the following: *logexper, mar*, *exper*.

C. Restricting your sample period to 1981 - 1987, estimate the union membership model using the **correlated random effects probit** estimator. Again, you should include occupational dummies but exclude *lagged union status.* Find the average marginal effects of the following: *logexper*.

D. Using the dynamic random effects probit model specified in Section 15.8.4 on pages 625-630 of Wooldridge's (2010) text, estimate the union membership model.[4] Of course, you now include *lagged union status* as a regressor! Test for state dependence in unionism. Compare your regression results with those reported in the last two columns of Table II of Vella and Verbeek (1998). (Note that the estimator you are using and the one used by the authors are somewhat different.)

Remarks - Please follow my general guidelines in reporting regression results. In particular, for each empirical question, include summary table(s), program used, and well-labeled output.

[4]Note that Wooldridge's suggestion is to use the standard probit random effects routine, but you expand the list of explanatory variables to include the initial state variable (y_0 = union membership status at the beginning of the study) and per period variables (z_i using Wooldridge's notation) generated from TVCs in the model. In this exercise, you can use overtime means of TVCs rather than per period variables. So, proceed as follows. (i) Generate the initial state variable, union membership status in 1980, say *union1980*; (ii) generate overtime means for all TVC variables (see my Illustration #5 under estimation of Chamberlain's RE Probit on how such variable can be generated in Stata); (iii) use Stata's RE probit routine: include all regressors in Table II of Vella and Verbeek plus *union1980* as well as all overtime means of time-varying regressors. Remember that, for time-varying covariates, the within variation is positive (e.g., within Std. Dev. > 0 from the *xtsum* command in Stata). Your estimation period for the dynamic model is 1981 to 1987 (3815 observations).