Panel Data Analysis of Microeconomic Decisions
Part II Assignment

Due Date: December 6 2024

Assignment Instructions

- Your assignment should include the relevant Stata code and output, as well as your answers to the questions (maximum 10 pages in total).
- Note that **relevant Stata output should be in the text (not in the Appendix)** and counts towards the page limit.
- **Code should be added at the end of the assignment** and does not count towards the page limit.
- Each student must submit individual answers that differ from the answers of the other students.
- Please submit your assignment on Canvas. If you have any questions, please send them by email to w.chen@uvt.nl.

Data

We use a balanced panel of five waves of data on men and women in a household with at least one person of age 50-61 in wave 1 who are responsible for their household's financial matters, taken from the US Health and Retirement Study (HRS). The waves were collected in 1994, 1996, 1998, 2000 and 2002. Respondents reporting zero financial wealth or with missing information on ownership of stocks or on the value of their stocks are deleted from the data, leading to a balanced panel of 11040 observations on 2208 individuals.

The variables that can be used are the following:

Variable Name	Description
ident	Person identifier in HRS
wave	Wave; 1=1994, ..., 5=2002
age	Age
marrlt	Dummy for married or living together
hispanic	Dummy for Hispanics
edyrs	Years of education
nonwhite	Dummy non-white
female	Dummy for females (0 for men, 1 for women)
totfw	Total financial wealth (in $ of 2002)
share	Share of total financial wealth invested in stocks (including stock mutual funds)

Table 1: Variable list

- The variables are stored in the Stata file stocks_balanced.dta. In the exercises below, we will use two dependent variables. The first is a dummy for owning stocks, i.e., for $share > 0$. The second is *share* itself. The other variables can be used as independent variables.

- Students will work on different samples. Sort the data on *ident*. The sample you have to use depends on the last digit of your student number (SNR): if the last digit is Y, you use the households $110Y + 1$ until $110Y + 1218$ ($Y = 0, 1, \ldots 9$) in your sorted dataset. (Check that this is a balanced panel with 1218 individuals and 6090 observations.)

Layout/presentation/quality of the text

The total possible points are 100, and up to 10 points may be deducted for issues with layout, presentation, or overall quality of the text.

Question 1 (5 points)

1. Select the data set you are supposed to use. Check this carefully! Make sure you understand the meaning of all the variables. Use *xtset* to make Stata recognize the panel structure of the data.(0.5 points)

2. Construct the dummy variable for ownership of stocks and stock mutual funds *dstocks* and compute the ownership rate for each of the five waves for your sample.(1 point)

3. Compute the 25th, 50th and 75th percentiles of the distribution of *share* for each wave. What do you conclude? (3.5 points)

Question 2 (25 points)

1. Estimate a static random effects probit model explaining *dstocks* from *female*, *age*, *edyrs*, *nonwhite*, *hispanic*, *marrlt* and *ltotfw* where *ltotfw* is the logarithm of *totfw*. Briefly interpret what you consider the main results.(5 points)

2. What can you say about the importance of unobserved heterogeneity according to this model?(5 points)

3. Construct dummies for each wave (except 2002) and add them to the model. Check whether they are jointly significant.(5 points)

4. Carry out some formal test to choose between three models: the model of (2.1) without any time effects, the model of (2.3) with time effects, or an intermediate model that includes a linear time trend (but no time dummies). (5 points)

5. Estimate the logit analogue of your favorite model in the previous question. What do you prefer here, logit or probit, why? Motivate your answer. (5 points)

Question 3 (25 points)

1. Estimate a dynamic random effects probit model explaining *dstocks* from the same variables as in Exercise 2 and from the first-order lag of *dstock*. Follow the Wooldridge (2005) approach to account for the initial condition problem. (Hint: first construct the initial values *share_0* and *dstock_0*, using, for example, the lag operator in Stata.)(5 points)

2. What can you say about the importance of the individual effect according to this model, in comparison to the model of Exercise 2? (5 points)

3. Test for the presence of state dependence. (5 points)

4. Compute the marginal effect of *lag_dstock* for the household with probability of owning stocks equal to the ownership rate in the complete sample.(5 points)

5. Test whether the individual effects are correlated with the initial value of *dstock*.(5 points)

Question 4 (25 points)

1. Estimate a static random effects tobit model explaining *share* from *female*, *age*, *edyrs*, *nonwhite*, *hispanic*, *marrlt* and *ltotfw*. Briefly interpret what you consider the main results. Note: don't forget the option ",ll(0)". (5 points)

2. Estimate the variance of the complete individual effects, using the output of this model. What can you say about the importance of unobserved heterogeneity? Compute the marginal effect of a change in *totfw* with 1% on the expected share invested in stocks, for someone with stock ownership probability 0.5. (5 points)

3. Test whether adding *ltotfw* squared is a significant improvement. (5 points)

4. In the model including both *ltotfw* and its square, compute the marginal effect of a change in *totfw* with 1% on the expected share invested in stocks, for someone with stock ownership probability 0.5 as a function of *ltotfw*. Draw a graph and interpret the result. (5 points)

5. Also try two-limit tobit, accounting for censoring of *share* at 1. Compare the results with those of 4.1. What do you conclude? Does this surprise you? (5 points)

Question 5 (20 points)

Your colleague claims that the individual effects (that is the parts independent of the initial outcome) are probably correlated with *ltotfw*, since wealthy people have different risk attitudes than less wealthy people.

1. Extend the model estimated in Question 4 to account for the possibility of such a correlation. You can assume that the relation between the individual effect and wealth remains constant over time. Estimate the extended model. (Hint: use the stata command "egen ...., by(....)" to construct an additional regressor.) (5 points)

2. Test whether the individual effects are correlated to *ltotfw* using the results in the previous question. (5 points)

3. Compare the coefficient on *ltotfw* with the estimate obtained in Exercise 4. Interpret the result.(5 points)

4. Estimate the variance of the ***total*** individual effects as a fraction of the variance of the total individual effects combined with the error terms according to the model estimated in 1 (ignoring the result of question 2). Interpret the result. (5 points)