FIE401 - Third assignment
Does financial literacy increase participation in the stock market?

Dmitrii Pugachev

Overview

Individuals have become increasingly active in financial markets. This increase in the market participation by individuals has been accompanied or even promoted by the advent of new financial products and services (e.g., commission-free trading offered by Robinhood). Some of the introduced products are complex and difficult to grasp, especially for financially unsophisticated investors (e.g., different class structures of mutual funds).

At the same time, market liberalization and structural reforms to social security and pensions have caused an ongoing shift in decision-making responsibility away from the government and employers and towards private individuals. Thus, individuals have to take more responsibility for their own financial well-being. This process is not universally seen as positive since many individuals have no skills and knowledge of how to do it.

This started a long string of research on the stock market participation of private individuals. In this exercise, you investigate one of the first questions asked in this literature: Are individuals with more finance related knowledge (a concept often referred to as financial literacy) more likely to participate in the stock market?

Formalities

Work together in assigned groups (check "People" section in the course navigation panel on Canvas, then "Groups" tab; search for your name; find the "G* Assignment 3" group where * is a number).

Deadline for assignment's submission is November 2 at 14:00. Submit your assignment even if you do not finish all tasks. To get a pass, 50% has to be correct.

Please submit two files:

- your **commented** coding file (file extension: ".R")
- a report with the numerical results and answers to assignment's questions (file extension: ".pdf")

Solution to the assignment will be released on November 2 at 17:00.

Submissions on November 2 after 14:00 but before 17:00 will get a warning. Two warnings or one submission after 17:00 - no access to the exam.

Please comment your code neatly so that a reader can reconstruct your thinking. You need neither to explain the used functions nor to describe your code in the report. Please keep your answers brief.

On Canvas, you will find *improved* guidelines of how to make good tables. Note that we will cover the subject of how to present research in detail in the last lecture of this course.

After submission, the assignments will be randomly redistributed. Read your peers' work carefully and compare your solution to theirs. There are many different approaches to code the same exercise, so let's learn from each other.

To make it easy for us to allocate individual assignment submissions, please follow the suggested file name structure:

- For the code: "Assignment3_X.R"
- For the report: "Assignment3_X.pdf"

where "X" indicates the group that submitted the assignment (e.g., "G2" or "G15").

Please do not include any personal information in the submitted files.

Please use online resources to clarify unfamiliar concepts.

If you are unsure how to use a certain command, either use R's own documentation (type `?command_name()`) or use *www.stackoverflow.com*.

Data provided for this assignment

The `Finlit.csv` file provides a dataset allowing you to explore the aformentioned question. In specific, you may find following variables.

- Main dependent variable
  - $mkt.part$: dummy for stock market participation
- Main independent variable
  - $adv.lit.index$: advanced literacy index
- Available control variables:
  - age: age bracket
  - $edu2$: dummy for pre-vocational education
  - $edu3$: dummy for intermediate vocational education
  - $edu4$: dummy for secondary pre-university education
  - $edu5$: dummy for higher vocational education
  - $edu6$: dummy for university education
  - $male$: dummy for gender
  - $partner$: married or not
  - $numkids$: number of kids
  - $retired$: retired or not
  - $selfempl$: dummy for self-employment
  - $lincome$: log(household income)
  - $nonequity.wealth.cat$: wealth quartile
  - $bsc.lit.index$: basic literacy index
  - $b2$: economics education
  - $b3$: daily use of economics
- Potential instruments
  - $f10$: sibling's financial situation
  - $f15$: parent's financial knowledge
- Included variables not relevant for current analysis
  - $edu1$: dummy for primary education
  - $b1$: self-assessed literacy

Part 1: LPM and instruments

You will replicate some results from Van Rooij et al (2011) in this part.

- Prepare a *Regression table* with the following three models:
  - Linear probability model

    $$mkt.part = \alpha_0 + \beta_1\ lit.index + \beta_X\ Controls + u$$

    where *Controls* indicate the list of control variables. Your goal is to replicate the second model of Table 7 (page 461). Include the same controls as in the table.
  - First stage of the IV regression

    Use four binary variables as instruments: (1) sibling's financial situation is "worse", (2) sibling's financial situation is "better", (3) parent's financial knowledge is "intermediate or high", and (4) parent's financial knowledge is "don't know". Your goal is to replicate the second model of Table 8A (page 463). Include the same controls as in the table.
  - Second stage of the IV regression estimated in one go

    Your goal is to replicate the second model of Table 8B (page 464). Include the same controls as in the table.
- Interpret your results. What does financial literacy mean for stock market participation, statistically and economically? Is there a noticeable difference between LPM and IV estimates?
- Assess relevance and exogeneity of the instruments. Are they good from your point of view?
- Carefully read chapter 5 (start from page 461 at the second paragraph beginning with "In Table 7,. . . ") of the paper and answer the following questions:
  - What reasons do the authors name for using an Instrumental Variable approach? Do you agree with these reasons?
  - After reading the authors' explanations and assessing the results, are you convinced that both instruments are good?

Comments

- Make sure that your pdf starts with a **short and consise** abstract presenting your analysis (as you would in your Master thesis).
- It is fine if your regression coefficients do not coincide but are comparable to coefficients in Van Rooij et al (2011).
- Report t-statistics rather than standard errors. Correct for heteroscedasticity of standard errors.
- In writing, discuss your regression results briefly (5-10 sentences). Make sure that you discuss statistical as well as economic significance. Regarding the latter, argue whether or not the detected effects are meaningful.
- The report has to be self-contained. In other words, it needs to contain all tables and any other piece of analysis that you reference in writing.

Grading guidelines

You can score 100 points for the assignment:

- Abstract [15 points]
- Regression table [50 points]
  - Linear probability model [20 points]
  - First stage of the IV regression [15 points]
  - Second stage of the IV regression [15 points]
- Interpretation of the results [15 points]
  - Statistical significance [5 points]
  - Economic significance [5 points]
  - Comparison of OLS and IV estimates [5 points]
- Assessment of the instruments [10 points]
  - Relevance [5 points]
  - Exogeneity [5 points]
- Questions based on chapter 5 of Van Rooij et al (2011) [10 points]
  - Reasons for IV [5 points]
  - Personal opinion [5 points]

Penalties

- Abstract has more than 150 words [-10 points]
- Report is hard to read (spelling / grammatical errors, unnatural language) [up to -15 points]
- Regression table is copy-pasted as a screenshot [-15 points]
- No correction of standard errors or t-statistics not reported [up to -5 points per regression]

Additional penalties are based on the relative quality of works.

References

- Van Rooij, M., Lusardi, A., and Alessie, R., "Financial Literacy and Stock Market Participation", 2011, Journal of Financial Economics, 101, 449-472.