Assignment 3 – FIE401

# Table of contents

Ds

# Abstract

Ds

# Regression table

This section delves into a nuanced discussion based on a series of regression models in a Stargazer table in the appendix. The analysis encompasses three main models: the Linear Probability Model (LPM), the First Stage Instrumental Variable (IV) Regression, and the Second Stage IV Regression.

## Linear Probability Model (LPM)

We initiate our analysis with the Linear Probability Model, focusing on the association between market participation and advanced literacy. Our modelling strategy closely aligns with the second model in Table 7 on page 461 of the referenced academic work. This is crucial for including a comprehensive range of demographic and financial control variables like age, education, and income. Certain continuous and ordinal variables have been converted to binary to streamline computations. The coefficient of 0.089 for the Advanced Literacy Index is statistically significant at the 1% level, indicating a robust and highly reliable positive association with market participation.

Moreover, control variables such as a log of income, wealth quartile 2, and wealth quartile 4 are also statistically significant. With an adjusted *R*2 of 0.120, the model accounts for 12% of the variance in market participation. However, it is worth noting that the LPM is susceptible to endogeneity, setting the stage for further analysis via IV models.

## First Stage IV Regression

Transitioning to the First Stage of our IV regression, we introduce four instrumental variables connected with the individual’s family’s financial conditions and knowledge. These instruments are congruent with the second model in Table 8A on page 463 of the referenced study, thereby ensuring methodological continuity. An array of control variables resembling those in the LPM are also incorporated. The variables indicating siblings with lesser financial knowledge and parents with unknown financial status stand out as highly significant instrumental variables, bolstering their credibility and effectiveness in the model. The variables for secondary pre-university education, higher vocational education, university education, and male also hold statistical significance among the controls. The adjusted *R*2 improves to 0.206, signifying a superior model fit compared to the LPM.

## Second Stage IV Regression (One-Go Estimation)

Lastly, the second stage of the IV regression is conducted in a single pass to extract causal inferences between the advanced literacy index and market participation. The selection of instruments and controls is harmonious with earlier models, ensuring analytical consistency. The coefficient of the Advanced Literacy Index elevates to 0.151 and is significant at the 5% level, further cementing its positive influence on market participation. Control variables like log of income, wealth quartile 3, and wealth quartile 4 maintain their significance, whereas the adjusted *R*2 marginally declines to 10.4%, which can be attributed to the model’s inherent complexity.

There is a noticeable difference between the LPM and IV estimates. While the LPM provides a good baseline understanding, the IV models go further by accounting for endogeneity, offering more reliable, causal insights. The coefficient in the IV model is higher, suggesting that when we control for endogeneity, the impact of financial literacy on stock market participation is even more pronounced than initially estimated by the LPM.

To sum up, our empirical analysis robustly validates the positive relationship between advanced literacy and market participation. The Instrumental Variable approach substantiates these findings, adjusting for potential endogeneity. Although some controls manifest varying degrees of significance across the models, they don't dilute the primary outcome: enhanced financial literacy positively correlates with active market participation.

## Financial literacy on market participation

Financial literacy refers to the ability to understand and use different kinds of financial tools. It is likely that people with a higher level of financial literacy use less income, create a retirement fund, and make suitable investments compared with people with a lower level of financial literacy. As we can see from our regression table the coefficient of advanced literacy index is significant at a 1 % level in the standard OLS model, and significant at a 5 % level in the IV model. The value of the coefficient is respectively 0.089 in the standard OLS model and 0.151 in the IV model, and we can therefore conclude that it excites a positive association between the market participation and higher level of financial literacy. If advanced literacy index increases by one, the probability that the person participate in the stock market increases by 0.089 due to the OLS model. Using our statistics from our regression table it is likely that people with increased level of financial literacy are better suited to conduct informed investment decisions and understand the risk and rewards with the stock market.

If we look at financial literacy in an economic perspective, it can lead to better financial outcomes by elevating their knowledge and be more financial literate. From an economic perspective it is more likely that a financial literate investor knows concepts as diversification, make decisions based on their risk tolerance and are able to navigate in the complex world of financial markets. This has policy implications: improving financial education could likely lead to increased stock market participation, thereby promoting economic stability.

Ds

Table 7:

A screenshot of a computer

Description automatically generated

Table 8A:

A table of numbers with black text

Description automatically generated with medium confidence

Table 8B:

A screenshot of a computer

Description automatically generated

# Interpretation of results

Ds

**“What does financial literacy mean for stock market participation, statistically and economically? Is there a noticeable difference between LPM and IV estimates?”**

Financial literacy refers to the ability to understand and use different kinds of financial tools. It is likely that one with a higher level of financial literacy use less income, create a retirement fund and make good investments compared with people with a lower level of financial literacy. As we can see from our regression table the coefficient of advanced literacy index is significant at a 1 % level in the standard OLS model, and significant at a 5 % level in the IV model. The values of the coefficient is respectively 0.089 in the standard OLS model and 0.151 in the IV model, and we can therefore conclude that it excites a positive association between the market participation and higher level of financial literacy. If advanced literacy index increases by one, the probability that the person participate in the stock market increases by 0.089 due to the OLS model. Using our statistics from our regression table it is likely that people with increased level of financial literacy are better suited to conduct informed investment decisions and understand the risk and rewards with the stock market.

If we look at financial literacy in an economic perspective, it can lead to better financial outcomes by increasing their knowledge and be more financial literate. From an economic perspective it is more likely that a financial literate investor knows concepts as diversification, make decisions based on their risk tolerance and are able to navigate in the complex world of financial markets. This has policy implications: improving financial education could likely lead to increased stock market participation, thereby promoting economic stability.

# Assessment of relevance and exogeneity of the instruments

Are they good from your point of view?

A crucial component of Instrumental Variables Regression (IVR) is checking the instrument validity. This means assessing the relevance and exogeneity of the instruments, which are the two assumptions for a valid use of the IV method.

The instruments are **exogenous** if all instruments are uncorrelated with the error term.

Let’s assess the relevance of the instruments (1):

The instruments are **relevant** if a least one instrument must enter the population counterpart of the first stage regression. Another interpretation of this is that at least one of the instruments must be related to our endogenous variable (x-variable, maybe say the name?) to be useful in addressing endogeneity. Essentially it refers to the strength of the relationship between the instruments and the endogenous variable. Weak instruments

1. \*\*Relevance\*\*:

Relevance refers to the strength of the relationship between the instruments and the endogenous variable(s). Instruments should be correlated with the endogenous variable(s) to be useful in addressing endogeneity.

a. \*\*F-Statistic\*\*: Calculate the first-stage F-statistic. It tests whether the instruments are jointly statistically significant in explaining the endogenous variable. A high F-statistic indicates relevance.

b. \*\*Partial F-Statistics\*\*: You can also calculate partial F-statistics for each instrument separately to assess their individual relevance.

c. \*\*Graphical Examination\*\*: Create scatter plots or partial regression plots to visually assess the relationship between each instrument and the endogenous variable.

2. \*\*Exogeneity\*\*:

Exogeneity refers to the assumption that the instruments are uncorrelated with the error term in the structural equation.

a. \*\*Over-Identification Test\*\*: Conduct an over-identification test, such as the Sargan-Hansen test. This test assesses the validity of the exogeneity assumption. A non-rejection of the null hypothesis suggests exogeneity.

b. \*\*Hausman Test\*\*: Compare the IV estimator to the OLS estimator. A non-rejection of the null hypothesis in a Hausman test indicates that the IV estimator is consistent, suggesting exogeneity.

c. \*\*Theoretical Considerations\*\*: Consider the theoretical underpinnings of your model. If the instruments are derived from variables that are exogenous by nature, it provides some evidence of exogeneity.

d. \*\*Instrumental Variable Tests\*\*: There are various diagnostic tests, such as the Durbin-Wu-Hausman test, which can be used to assess the exogeneity of specific instruments.

It's important to remember that assessing the relevance and exogeneity of instruments in IV regression can be complex and may involve several tests and considerations. In practice, you should carefully examine the specific context of your study, the nature of your instruments, and the validity of the underlying assumptions.

Additionally, it's often a good practice to report the results of these assessments in your research paper or report, along with the potential limitations and robustness checks to enhance the credibility of your IV regression results.

# Additional questions

## What reasons do the authors name for using an Instrumental Variable approach? Do you agree with these reasons?

ds

## After reading the authors’ explanations and assessing the results, are you convinced that both instruments are good?

ds

# Appendix

## Regresssion table

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  | Dependent variable: | | |
|  |  | | |
|  | Linear probability model | First stage IV | Second stage IV in one go |
|  | OLS | OLS | instrumental |
|  |  |  | variable |
|  | (1) | (2) | (3) |
|  | | | |
| Advanced literacy index | 0.089\*\*\* |  | 0.151\*\* |
|  | (0.012) |  | (0.060) |
|  |  |  |  |
| Sibling less financial knowledge |  | 0.425\*\*\* |  |
|  |  | (0.082) |  |
|  |  |  |  |
| Sibling better financial knowledge |  | 0.176\*\* |  |
|  |  | (0.077) |  |
|  |  |  |  |
| Parents intermediate or high financial knowledge |  | -0.205\*\* |  |
|  |  | (0.088) |  |
|  |  |  |  |
| Parents unknown financial knowledge |  | -0.727\*\*\* |  |
|  |  | (0.152) |  |
|  |  |  |  |
| Age 30-40 | 0.011 | -0.061 | 0.016 |
|  | (0.047) | (0.131) | (0.048) |
|  |  |  |  |
| Age 40-50 | 0.040 | -0.100 | 0.047 |
|  | (0.049) | (0.128) | (0.050) |
|  |  |  |  |
| Age 50-60 | 0.027 | 0.021 | 0.026 |
|  | (0.049) | (0.129) | (0.049) |
|  |  |  |  |
| Age over 60 | 0.071 | 0.002 | 0.072 |
|  | (0.060) | (0.149) | (0.059) |
|  |  |  |  |
| Intermidiate vocational education | 0.023 | 0.096 | 0.016 |
|  | (0.037) | (0.097) | (0.037) |
|  |  |  |  |
| Secondary pre-university education | 0.028 | 0.495\*\*\* | -0.004 |
|  | (0.042) | (0.091) | (0.053) |
|  |  |  |  |
| Higher vocational education | 0.071\* | 0.391\*\*\* | 0.045 |
|  | (0.037) | (0.078) | (0.045) |
|  |  |  |  |
| University education | 0.102\*\* | 0.525\*\*\* | 0.066 |
|  | (0.048) | (0.087) | (0.060) |
|  |  |  |  |
| Male | 0.070\*\* | 0.441\*\*\* | 0.044 |
|  | (0.028) | (0.061) | (0.036) |
|  |  |  |  |
| Partner | 0.026 | 0.113 | 0.019 |
|  | (0.032) | (0.070) | (0.032) |
|  |  |  |  |
| Number of children | 0.002 | -0.040 | 0.005 |
|  | (0.015) | (0.033) | (0.015) |
|  |  |  |  |
| Retired | -0.021 | -0.015 | -0.021 |
|  | (0.052) | (0.101) | (0.050) |
|  |  |  |  |
| Self-employed | 0.032 | 0.145 | 0.022 |
|  | (0.059) | (0.096) | (0.059) |
|  |  |  |  |
| Log of income | 0.085\*\*\* | 0.033 | 0.080\*\*\* |
|  | (0.027) | (0.062) | (0.027) |
|  |  |  |  |
| Wealth quartile 2 | 0.075\*\* | 0.277\*\*\* | 0.058 |
|  | (0.035) | (0.102) | (0.040) |
|  |  |  |  |
| Wealth quartile 3 | 0.116\*\*\* | 0.387\*\*\* | 0.090\*\* |
|  | (0.037) | (0.097) | (0.045) |
|  |  |  |  |
| Wealth quartile 4 | 0.161\*\*\* | 0.524\*\*\* | 0.126\*\* |
|  | (0.043) | (0.105) | (0.056) |
|  |  |  |  |
| Constant | -0.796\*\*\* | -1.085\* | -0.696\*\* |
|  | (0.264) | (0.612) | (0.271) |
|  |  |  |  |
|  | | | |
| Observations | 1,115 | 1,115 | 1,115 |
| R2 | 0.134 | 0.221 | 0.119 |
| Adjusted R2 | 0.120 | 0.206 | 0.104 |
| Residual Std. Error | 0.425 (df = 1096) | 0.897 (df = 1093) | 0.429 (df = 1096) |
| F Statistic | 9.455\*\*\* (df = 18; 1096) | 14.756\*\*\* (df = 21; 1093) |  |
|  | | | |
| Note: | \*p\*\*p\*\*\*p<0.01 | | |