In this section we will reference to the regression table which consist of a stargazer table with the following outputs side by side. Due to the large table, we have inserted this in appendix.

## Linear probability model

We start by exploring the relationship between market participation and the advanced literacy index while controlling for several demographic and financial variables. These controls are integral to the robustness of our model and include categories such as age groups, levels of education, gender, marital status, number of dependents, retirement status, self-employment, linear income, and non-equity wealth. Our model aims to emulate the second model shown on Table 7, page 461, of the referenced study, incorporating the same control variables for consistency and comparative analysis.

We converted specific variables like age and non-equity wealth into binary forms to facilitate the computational aspects. This data manipulation allows us to seamlessly incorporate these variables into our linear probability model.

Through this comprehensive approach, our model not only examines the relationship between financial literacy and market participation but does so while taking into account multiple influencing factors. This ensures that the insights gleaned are as accurate as possible, thus providing a solid foundation for any subsequent interpretation or policy recommendations.

The coefficient of the Advanced literacy index is signifiacnyt at 1% level, suggesting that an increase in advanced literacy is positively related to market participation. Among the controls, **lincome**, **wq3**, and **wq4** are significant, implying that these factors also influence market participation. The adjusted �2*R*2 is 0.120, meaning the model explains 12% of the variability in market participation. While the OLS model provides a baseline understanding, it may suffer from endogeneity, potentially biasing the results. Therefore, we proceed to the IV models for a more rigorous analysis.

## First stage IV

In the first stage of the instrumental variable (IV) regression, we employ four binary variables as instruments. These variables relate to the financial situation and knowledge of the individual's family, specifically their siblings and parents. The variables are as follows: 1) the financial situation of the individual's sibling being "worse," 2) the financial situation of the sibling being "better," 3) the parent’s level of financial knowledge being "intermediate or high," and 4) the parent’s level of financial knowledge being "don't know."

The selection of these instrumental variables is purposeful, aiming to mirror the methodology laid out in the second model of Table 8A on page 463 of the referenced literature. This ensures comparability and robustness in the analysis.

As in the linear probability model, we incorporate a wide range of control variables in this first stage. These include age groups, educational attainment, gender, marital status, number of dependents, retirement status, self-employment, linear income, and non-equity wealth categories.

We converted certain ordinal and nominal variables into binary format for computational efficiency, just as we did in the linear probability model. This ensures a smoother inclusion of these variables in the model and facilitates a more straightforward interpretation of the results.

In summary, the first stage of our IV regression is designed to rigorously examine the relationships affecting the advanced literacy index, using a comprehensive set of instruments and control variables. The model provides essential groundwork for the second-stage IV regression that follows.

From the stargazer output we can see that instrumental variables siblings\_worse and parents\_dont\_know have high signinifcance levels, indicating that they are strong instruments. **sibling\_better** and **parent\_intermediate\_high** are also statistically significant, albeit at lower levels. For the control variables are the educational levels edu4, edu 5 and edu 6 are statistically significant, as is male. he adjusted �2*R*2 rises to 0.206, suggesting that the first stage model fits the data better than the OLS model. The First Stage IV model shows that our chosen instruments are valid and strong, indicating a good fit for our IV estimation strategy. The controls largely maintain their significance.

## Second stage IV in one go

In the final step of our econometric analysis, we execute the second stage of the Instrumental Variable (IV) regression in a single pass. At this stage, the primary objective is to discern the causal relationship between market participation and the advanced literacy index. This "one-go" estimation integrates both stages of the IV regression, optimizing computational efficiency and analytical precision.

To align with the referenced academic framework, the model has a set of controls identical to those employed in earlier stages. For the instrumental variables, we retain the four binary variables connected to the financial condition and knowledge of the individual's siblings and parents, as formulated in the first stage. These instruments are conjoined with the full range of control variables used in the previous models, offering a robust and comprehensive framework for causal inference.

This unified estimation approach allows us to derive more accurate and unbiased coefficient estimates, showing how the advanced literacy index impacts market participation while controlling for potential endogeneity.

In sum, the second stage of the IV regression, estimated in one go, is the culmination of our econometric inquiry, providing us with the most unbiased estimates to understand the underlying causal relationships.

The coefficient jumps to 0.151 and is significant at the 5% level, confirming the positive relationship between advanced literacy and market participation. Some variables like **lincome**, **wq3**, and **wq4** remain significant, while the educational variables lose their significance. The adjusted �2*R*2 slightly decreases to 10.4%, but this is expected given the complexity of the IV model. The Second Stage IV model confirms the positive impact of advanced literacy on market participation, even after controlling for endogeneity. While some controls lose their significance, key variables like **lincome** maintain their importance.

## Conclusion

Our econometric models demonstrate a strong and positive relationship between advanced literacy and market participation. The Instrumental Variable approach confirms this finding while accounting for potential endogeneity. This lends robustness to our key hypothesis that improved literacy enhances market engagement. The controls offer additional insights but their impact varies across the different models.