

Waseda University

School of Political Science and Economics

Research project

Economic Policy 4

Quantitative Analysis, Fall 2024

Table of contents

Introduction	3
Research question	3
Hypotheses	3
Data	4
Variables	4
Descriptive statistics	5
Data Visualization	6
Regression Analysis	7
Model specification	8
Interpretation of the results	8
Adding control variables	9
Interpretation with control variables	10
Model Comparison	10
Conclusion	11
Appendix	12
References	12

Introduction

Introduction to the research project here.

Research question

How does objective financial literacy (as measured by quiz-style questions on interest, inflation, and compounding) relate to young adults' annual household income and willingness to take financial risks in the United States, and does ethnicity moderate these relationships?

Hypotheses

H1 (Income Relationship): Young adults in the U.S. who demonstrate higher objective financial literacy will report higher annual household income.

H2 (Risk Relationship): Young adults in the U.S. who demonstrate higher objective financial literacy will exhibit a greater willingness to take financial risks.

H3 (Ethnicity as Moderator): The positive relationships in H1 and H2 will differ by ethnicity, such that the effect of financial literacy on income and risk tolerance is stronger (or weaker) for certain ethnic groups.

Data

Explanation for the data here and what it is.

Variables

This needs to be written better:

We are going to look at the variables Objective Measure: M6, M7, M8, and M31 test concrete knowledge and are widely recognized as a legitimate indicator of financial literacy which will be the independent variable.

The dependent variables will be the annual household income and the willingness to take financial risks.

These three variables in the dataset are the following:

M6, M7, M8, and M31 (Objective Measure: M6, M7, M8, and M31 test concrete knowledge and are widely recognized as a legitimate indicator of financial literacy)

A8_2021 (Annual household income)

J2, which is risk tolerance.

A4A_new_w , if we also include ethnicity this will be the moderator variable.

- A50B (Binary variable for age and gender)

We also might control for these variables:

- A41 (Highest education of the person(s) that raised the respondent)
- P52 (Do you expect to inherit \$10,000 or more in the future?)

The variables going to be used in this analysis can be seen in the following [Table 1](#).

Variable	Description
Financial Literacy	Sum of 4 financial literacy questions, score from 0-4
Annual Household Income	From less than 15k to 300.000k
Risk Tolerance	Willingness to take financial risks
Ethnicity	Binary variable between white non-Hispanic and non-white
Sex	Binary variable between Male and Female
Education Parent	Highest education of the person(s) that raised the respondent
Inheritance	No or Yes between expected future inheritance

Table 1: Variables used in the analysis

To furthermore explain the variables, financial literacy is a sum of 4 financial literacy questions, with a score from 0 to 4. The annual household income is the annual household income with variables between 0-10 where 1 is less than 15,000 dollars and 10 is \$300,000 or more.

Risk tolerance is the willingness to take financial risks with variables between 1-10 where 1 is no risk at all and 10 is very willing to take risks.

Inheritance is a binary variable between no and yes, where no is no expected future inheritance and yes is expected future inheritance above 10,000 dollars.

Education parent is the highest education of the person(s) that raised the respondent, with variables between 1-6 where 1 is did not complete high school and 6 is a postgraduate

Ethnicity is a binary variable between white non-Hispanic and non-white.

Descriptive statistics

	Annual Household Income	Financial Literacy	Risk Tolerance
Mean	3.58	2.05	6.48
Std.Dev	2.19	1.11	2.57
Min	1.00	0.00	1.00
Q1	2.00	1.00	5.00
Median	3.00	2.00	7.00
Q3	5.00	3.00	8.00
Max	10.00	4.00	10.00
MAD	2.97	1.48	2.97
IQR	3.00	2.00	3.00
CV	0.61	0.54	0.40
Skewness	0.58	0.01	-0.28
SE.Skewness	0.09	0.09	0.09
Kurtosis	-0.38	-0.72	-0.83
N.Valid	784.00	784.00	784.00
Pct.Valid	100.00	100.00	100.00

Table 2: Descriptive Statistics for my data (N = 784)

The descriptive statistics for the data can be seen in [Table 2](#).

Need to finish descriptive statistics and explain the standard deviation, mean, and other statistics.

The descriptive statistics for annual household income show that the mean is 3.58, with a standard deviation of 2.19. The minimum is 1, which is less than 15,000 dollars, and the maximum is 10, which is 300,000 dollars or more. The median is 3, which is between 25,000 and 35,000 dollars. The interquartile range is 3, which is between 35,000 and 50,000 dollars. The skewness is 0.58, which is positive and means that the distribution right-skewed.

The kurtosis is -0.38, which means the distribution is flatter than the normal. The mean absolute deviation is 2.97, and the coefficient of variation is 0.61.

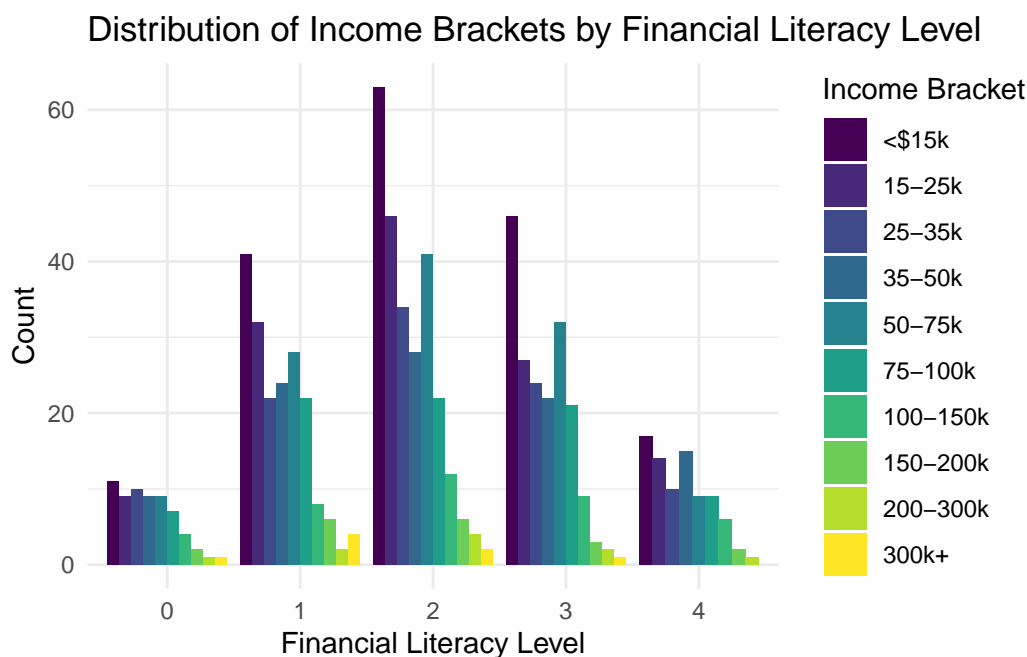
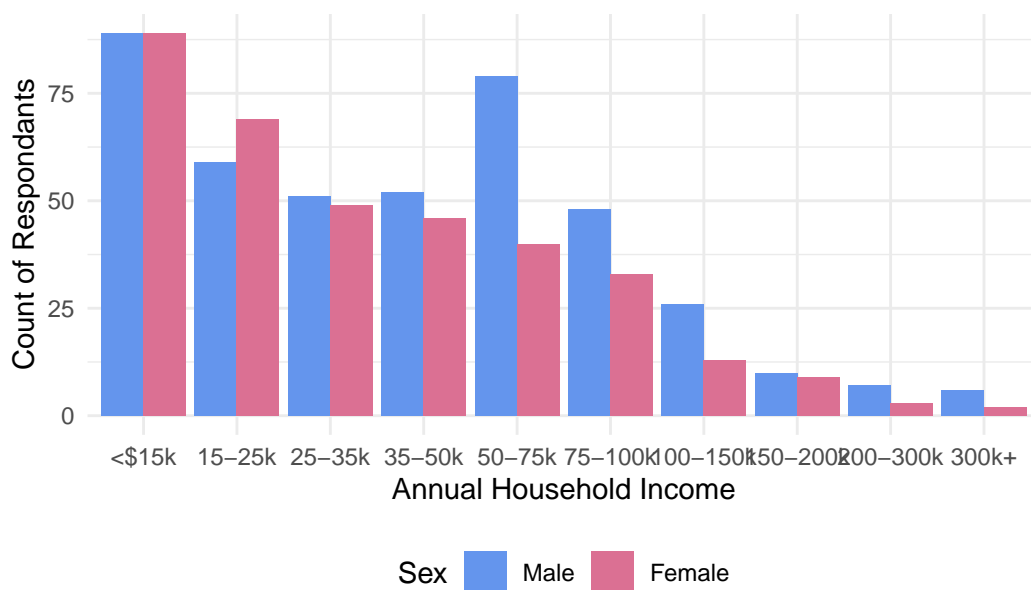
Financial literacy which is the sum of 4 financial literacy questions has a mean of 2.05, with a standard deviation of 1.11. The minimum is 0, and the maximum is 4. The

median is 2, and the interquartile range is 2. The skewness is 0.01, which means that the distribution is close to a normal distribution, and has no skewness or you can say it is symmetrical.

The the kurtosis is -0.72. A negative kurtosis means that the distribution is flatter than a normal distribution.

Data Visualization





Regression Analysis

Our group is going to use an ordinal logistic regression model to analyze the relationship between financial literacy, annual household income, and willingness to take financial risks in the United States for young adults. We will also investigate if financial literacy has a different effect on income and risk tolerance and moderate these relationships by sex, ethnicity (not sure yet), and expected future inheritance to see if the effect of financial literacy on income and risk tolerance is stronger (or weaker) for certain groups.

Model specification

The model we have to use is an ordinal logistic regression model, which is a type of regression analysis used for predicting the relationship between ordinal dependent variables and one or more independent variables. The dependent variable is ordinal, which means that it has a natural order, but the intervals between the categories are not equal. The independent variables can be either continuous or categorical.

The ordinal logistic regression model without control variables is specified as:

$$\text{logit}(P(\text{Income} \leq j)) = \alpha_j + \beta_1 \times \text{Financial Literacy}$$

where Annual Household Income is the dependent variable and Financial Literacy is the independent variable. A positive coefficient for Financial Literacy would indicate that higher financial literacy is associated with higher annual household income, or more precisely, a higher probability of being in a higher income category. In R we can fit this model using the `polr` function from the MASS package.

```
=====
                        Dependent variable:
-----
                        Annual_household_income
-----
Financial_literacy      -0.042
                        (0.056)
-----
Observations              780
=====
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Interpretation of the results

Model 1 with no control variables and interaction terms shows that the effect of financial literacy on annual household income is not statistically significant. The odds ratio coefficient for financial literacy is -0.042, which means that a one-unit increase in financial literacy is associated with a 4.2% decrease in the odds of being in a higher income category. Which is opposite to what we expected, but not statistically significant. This basically means that I am wasting my time studying economics and finance, I should have studied something else.

To explore this further, we will add control variables and interaction terms to the model.

Adding control variables

The model with control variables is specified as:

$$\text{logit}(P(\text{Income} \leq j)) = \alpha_j + \beta_1 \times \text{Financial Literacy} + \beta_2 \times \text{Ethnicity} + \beta_3 \times \text{Sex} + \beta_4 \times \text{Education Parent}$$

where we control for ethnicity, sex, education parent, and inheritance. We can fit this model using the `polr` function from the `MASS` package.

Dependent variable:	
Annual_household_income	
Financial_literacy	-0.135** (0.059)
EthnicityNon-White	-0.348** (0.175)
SexFemale	-0.862*** (0.186)
Education_parentHigh school	0.129 (0.318)
Education_parentSome college	0.650** (0.328)
Education_parentAssociate's degree	0.939*** (0.360)
Education_parentBachelor's degree	1.190*** (0.329)
Education_parentPost graduate degree	1.583*** (0.365)
EthnicityNon-White:SexFemale	0.750*** (0.256)
Observations	780
Note: *p<0.1; **p<0.05; ***p<0.01	

Interpretation with control variables

Model 2 with control variables and interaction terms shows that the effect of financial literacy on annual household income now becomes statistically significant at the 0.05 level. The odds ratio coefficient for financial literacy now becomes even more negative with a coefficient of -0.192, which means that a one-unit increase in financial literacy is associated with a 19.2% decrease in the odds of being in a higher income category. This is a very interesting result, and it suggests that financial literacy is negatively associated with annual household income.

We also find that being female is strongly associated with lower income, with a coefficient of -1.106. The same can be said for non-white, but to a lower effect where the coefficient is -0.371.

Maybe not so interestingly we find that the higher education of the person(s) that raised the respondent is also significantly associated with higher log odds of being in a higher income category. With 6.8% for Some college and up to 16.7% for a postgraduate degree.

Model Comparison

```
# Model comparison
anova(fit, fit2, test = "Chisq")
```

Likelihood ratio tests of ordinal regression models

Response: Annual_household_income

					Model	Resid.	df
1					Financial_literacy		770
2	Financial_literacy + Ethnicity + Sex * Ethnicity + Education_parent						762
	Resid. Dev	Test	Df	LR stat.	Pr(Chi)		
1	3155.128						
2	3074.197	1 vs 2	8	80.93154	3.175238e-14		

Doing an ANOVA test which is a test that compares the fit of two models to see if the more complex model is significantly better than the simpler model. This test here shows that model2 is not necessarily a better fit than model 1, and the residual deviance which is a measure of how well the model fits the data where lower is better is not significantly different between the two models. But the p value is extremely small, so we can reject the null hypothesis that the simpler model is better, which makes sense since we added multiple important control variables.

Conclusion

Overall the results show that financial literacy is negatively associated with annual household income, which is the opposite of what we expected. This is a very interesting result and suggests that financial literacy may not be as important as previously thought for young adults in the United States. We also find that being female and non-white is strongly associated with lower log odds of being in a lower income bracket, which is a concerning result. We also find that higher education is associated with higher income, which is a positive result.

Some limitations of this study are that after filtering the data for financial literacy, annual household income, and willingness to take financial risks, we were left with a relatively small sample size of 780 young adults in the United States out of 3009, where a lot got removed because they answered don't know or prefer not to answer. This could have biased the results, and we should be careful when generalizing the results to the entire population of young adults in the United States. Also, the sample size of non-white young adults was relatively small, which could have affected the results.

However this is just a correlation and not causation, so we cannot say that financial literacy causes lower income among young adults in the US, but it is an interesting result that should be further investigated in future research.

Appendix

References