

Waseda University

School of Political Science and Economics

Research project

Economic Policy 4

Quantitative Analysis, Fall 2024

Table of contents

Research question	3
Feedback	3
Finalized research question after feedback	3
Some different hypotheses	3
Data	4
Research feedback 2	4
Data after feedback	5
Data cleaning	5
Variables	8
Descriptive statistics	8

Research question

How does the level of financial literacy influence individual investment choices and portfolio diversification among young adults?

Feedback

I like this idea for the project, and I think it would work very well as an international comparison as you suggest, especially if you can find some specific cases of countries that have tried to implement education policies to improve financial literacy. My one concern is that while you say there is a reasonable amount of survey data, some of the ideas you suggest could be quite hard to operationalise - you need variables to measure the level of people's financial literacy, as well as their success in investment. Your most important task therefore is to find a clear, high-quality source for those variables. You should also think about what other factors might have an impact - for example, if you are doing a cross-national comparison, the availability of certain investment products may differ between countries, and currency trends could also impact people's potential investment returns, so it would be good to include that in your analysis.

In line with this feedback, your next step for this research should be to operationalise your questions and hypothesis, specifying exactly which variables you plan to use for your analysis and where you are going to get them from. This will ensure that you are in a good position to actually start gathering that data and putting it into R.

Finalized research question after feedback

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?

Some different 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

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.

Research feedback 2

Dear all,

(Economic Policy 4 Group)

This is great work - you've identified a really good dataset for your question and are obviously well on track to being able to start your analysis. Before you do so, there are a couple of things you might want to consider. Firstly, reported household income normally relates to salary - people often fail to include their investment returns when they report these numbers. That isn't a problem for your statistical analysis, but you might want to give some thought to how you will present the logic of your hypothesis; what is the causal mechanism by which you think financial literacy is connected to household income, if this often excludes investment income?

You may also want to have a look at your data set to see if there is any indicator that could suggest generational wealth (e.g., social class etc.) - there may not be anything, but generally speaking the strongest indicator of household income is the person's parents' household income, so controlling for that in some way would make your analysis of the effects of financial literacy much more robust. You may also want to include gender as well as ethnicity as a control variable, since we know that there is often an effect of gender on income opportunities.

Generally though I think you're very much on the right track with this research; before the winter break it would be great if you could get to the stage where you're starting to test some of these models in R, so that you can move on to interpreting results and generating graphs by the start of January.

Data after feedback

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, ethnicity will be the moderator variable.

New variables after feedback

- 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?)

Data cleaning

```
rm(list = ls())
library(tidyverse)
library(stargazer)
library(summarytools)
```

```
url <- "https://raw.githubusercontent.com/DanielFabioG/data/refs/heads/main/NFCS_2021_DA"
data <- read.csv(url)
```

```
# Selecting the variables of interest, renaming them,
# and filtering the data to include only young adults
data <- data %>%
  select(M6, M7, M8, M31, A8_2021, J2, A4A_new_w, A50B, A41, P52) %>%
  rename(Financial_literacy = c(M6, M7, M8, M31),
         Annual_household_income = A8_2021,
         Risk_tolerance = J2,
         Sex = A50B,
```

```

    Education_parent = A41,
    Inheritance = P52,
    Ethnicity = A4A_new_w) %>%
# 1 is male 18-24 while 7 is female 18-24
filter(Sex %in% c(1, 7))

# Making sex_and_age a male and female variable
data$Sex <- factor(data$Sex, levels = c(1, 7), labels = c("Male",
                                                         "Female"))

# Making ethnicity a factor variable
data$Ethnicity <- factor(data$Ethnicity, levels = c(1, 2),
                        labels = c("White non-Hispanic",
                                   "Non-White"))

# Fixing inheritance and making it a factor
data$Inheritance <- factor(data$Inheritance, levels = c(1, 2, 8, 9),
                          labels = c("Yes",
                                     "No",
                                     "Don't know",
                                     "Prefer not to answer"))

# Fixing education and making it a factor
data$Education_parent <- factor(data$Education_parent, levels = c(1:6),
                              labels = c("Did not complete high school",
                                         "High school",
                                         "Some college",
                                         "Associate's degree",
                                         "Bachelor's degree",
                                         "Post graduate degree"))

# Risk tolerance goes from 1 to 10, where 1 is not at all and 10 is very willing,
#we will just remove dont know and prefer not to answer
data <- data %>%
  filter(Risk_tolerance %in% c(1:10))

# Annual household income goes from 1 to 10 where 1 is less than $15,000
# and 10 is $300,000 or more
data <- data %>%
  filter(Annual_household_income %in% c(1:10))

# Financial literacy question 1:
# Suppose you had $100 in a savings account and the interest rate was 2%
# per year. After 5 years, how much do you think you would have in the
# account if you left the money to grow?

```

```
data <- data %>%
  filter(Financial_literacy1 %in% c(1:3))

# Financial literacy question 2:

#Imagine that the interest rate on your savings account was 1% per year
# and inflation was 2% per year. After 1 year, how much would you be able
# to buy with the money in this account?

data <- data %>%
  filter(Financial_literacy2 %in% c(1:3))

# Financial literacy question 3:

# If interest rates rise, what will typically happen to bond prices?

data <- data %>%
  filter(Financial_literacy3 %in% c(1:3))

# Financial literacy question 4:

#Suppose you owe $1,000 on a loan and the interest rate you are charged
# is 20% per year compounded annually. If you didn't pay anything off, at
# this interest rate, how many years would it take for the amount you owe
# to double?

data <- data %>%
  filter(Financial_literacy4 %in% c(1:3))

# Now to make the financial literacy variable we will recode
# the variables to be 1 if correct and 0 if incorrect
# The ifelse function takes the correct answer and assigns 1 to it,
#and 0 to the incorrect answers and stores it in the new variable
# where it will have a sum of 0-4
data <- data %>%
  mutate(Financial_literacy = ifelse(Financial_literacy1 == 1, 1, 0) +
         ifelse(Financial_literacy2 == 3, 1, 0) +
         ifelse(Financial_literacy3 == 2, 1, 0) +
         ifelse(Financial_literacy4 == 2, 1, 0))

# Removing the individual financial literacy questions
data <- data %>%
  select(-Financial_literacy1,
        -Financial_literacy2,
        -Financial_literacy3,
        -Financial_literacy4)
```

```
# Checking for missing values
data %>%
  summarise_all(~sum(is.na(.)))
```

```
Annual_household_income Risk_tolerance Ethnicity Sex Education_parent
1 0 0 0 0 4
Inheritance Financial_literacy
1 95 0
```

Variables

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	Annual household income
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

```
# Descriptive statistics
#descr(data, style = "rmarkdown")
```


	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)