Lam Tran

ECON 467

November 26, 2020

**Final Paper**

**Introduction**

In May 2020, the United States government passed the $2.2 trillion CARES Act that gave each qualifying individual a $1,200 check in response to the economic fallout as the consequence of the coronavirus. The check was supposed to allow people to keep paying their bills during the economic shutdowns, but other than using the check to pay bills, people in nearly every income bracket also spent it on trading stocks (Fitzgerald, 2020). To elaborate, people who earned more than $35,000 increased their stock trading by at least 50% compared to the week before they received the check. Even more than that, the check (along with the free time people had during lockdowns and the possibility of the market comeback after people “buy the dip”) allowed new investors to step into the market as well, as there was a surge in the number of new accounts opened in most major online brokers such as Charles Schwab, TD Ameritrade, Etrade, and especially, Robinhood (Fitzgerald, 2020).

Robinhood, a financial services company that was founded in 2013, was one of the pioneers who offered no-fee, no-minimum-balance stock, ETF, option, and crypto trading through its mobile app and website (Robinhood). Recently, the platform even offered fractional shares which would allow investors to buy a fraction of stock even with “as little as $1” (Robinhood). The company’s targeted customers are young and less well-off adults who were not being considered by other financial firms, thus the majority of Robinhood’s users are “millennials” who once felt investing was inaccessible due to fees and minimum account balance (Huang, 2015). The ability to check their trading accounts anytime and anywhere just from their phones is another plus that attracts these young, first-time investors. Although Robinhood remains the top suggested broker for beginners, after Robinhood was founded, other brokerage firms saw the potential in this customer base and created other no-fee, online-based trading platforms as well.

The phenomenon that occurred to the volume of traded stocks and the number of new trading accounts when stimulus checks were given raised an interesting hypothesis that without Robinhood and similar trading apps that lowered investing barriers, especially regarding income or the amount of available capital, maybe there would not have been that much participation in the stock market the moment people had a bit of free money. Therefore, this study is going to analyze if familiarity with mobile financial services would positively influence the likelihood of participating in the stock market. To be more specific, familiarity with mobile financial services could be measured by the frequency of using online banking, mobile banking, or mobile financial budgeting applications. If online financial services indeed have an economically significant impact on the stock market participation, they could be developed and used as another tool to encourage investing, the goal the Federal Reserve has aimed at through lowering the interest rate, especially when the interest rate is near 0 and could hardly be any lower.

There was one paper by Bogan (2008) concluded that Internet-using households participated in the stock market substantially more than households that did not use the Internet in the United States, which is in line with what this research is expected to find. However, the data the author used was up to the year 2002, when the Internet did not become as popular as today and no-fee, online-based financial services were not available, so the effect of the Internet will not necessarily remain the same compared to the present situation. Furthermore, Bogan’s data heavily sampled older people who tend to not use the Internet as much and thus would bias the results, and the research did not account for some key financial capability factors such as risk tolerance or financial literacy. Therefore, this research would improve upon Bogan’s research while examining whether people who are more familiar with online financial services tend to participate in the stock market more than people who are not, holding the aforementioned key factors constant. The cross-sectional data that will be used is from the State-to-State survey from the National Financial Capability Study that surveyed 27,091 American adults from 18 years old from June through October 2018 about their financial status and related financial capabilities.

This research would also build upon findings from other studies in other countries about stock market participation. These studies did not mention the effect of online financial services, but they had important conclusions about other key factors that affect stock market participation. One previous study has found out that there was a statistically significant relationship between financial literacy and stock market participation in the Netherlands (Rooij, Lusardi, and Alessie, 2011). To elaborate, those who scored lower on their financial knowledge test were much less likely to trade stocks or securities. Another study pointed out that there was a gender gap in stock market participation in Sweden, more specifically women were less likely to participate than men (Almenberg and Dreber, 2015). However, the authors also concluded that controlling for financial literacy would significantly reduce that gap to nearly nonexistent. Furthermore, the authors examined the risk tolerance between women and men and concluded that even when controlling for financial literacy, women were much more reluctant to take financial risks. As the two studies used data in recent years and first-world countries with similar economic outlooks and demographics, their conclusions are expected to stand true in this research’s model as well.

Consequently, this research was inspired to control for gender, risk tolerance level, and financial literacy level along with other key demographic factors (income, education level, age) to correctly quantify the effect of familiarity with online financial services on stock market participation. Nonetheless, it is expected that gender would not affect stock market participation after controlling for financial literacy as in the research by Almenberg and Dreber (2015).

**Data and Methods**

The data that will be used in this research is from the National Financial Capability Study (NFCS), which is a project of the FINRA Investor Education Foundation. The data is from a state-by-state survey that was conducted online from June through October 2018 among a nationally-representative sample of 27,091 American adults, reaching approximately 500 individuals per state plus the District of Columbia, except for Oregon and Washington with oversamples of 1,250. After removing missing observations for the variables that will be used in this research, there are 22,368 observations left. The missing observations spread across variables and there is no reason to suspect that remove them would create a systematic bias to the model. All numerical values are within fixed ranges so there is no outlier that could bias the models. However, the data was collected online, so the respondents could be more familiar with technology than the average population (sampling bias).

In the estimated model, other than the key independent variable which is familiarity level with financial mobile apps, other key factors that will be controlled for are important demographic and financial factors: gender, ethnicity, age, income, risk tolerance level, debt level, education level, self-assessed financial literacy level, and the financial literacy level based on the given financial quiz in the survey. As all variables that hypothetically could influence the likelihood to participate in the stock market are included in the model, the omitted variable bias should not be present. I will be using the Linear Probability Model to estimate whether familiarity with financial mobile apps has a positive impact on the likelihood to participate in the stock market.**Participation = B1\*Familiar+ B2\*Male + B3\*Income + B4\*Risk + B5\*Debt + B6\*LowTest + B7\*HighTest + B8\*White + B9\*LowEd + B10\*HighEd + B11\*FinLit**

Table 1: Model Variables and Estimated Signs (The excluded group is MidTest for people who correctly answer 3-5 questions, (The excluded group is MidEd for people who have finished high school/GED or have some college education)

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Predicted Sign | The Measurement |
| Participation | A categorical variable for participation in the stock market: | N/A | = 1 if participate  = 0 if not participate |
| White | A categorical variable for ethnicity: | + | = 1 if white  = 0 if non-white |
| Male | A categorical variable for gender | + | = 1 if male  = 0 if female |
| Age | A discrete variable for age given in years derived from the mean of the given age groups | + | = 21 if 18-25 years old  = 30 if 25-34 years old  = 40 if 35-44 years old  = 50 if 45-54 years old  = 60 if 55-64 years old  = 70 if 60+ years old |
| Income | A discrete variable for annual income in thousands of dollars derived from the mean of the given income brackets | + | = 10 if < $15,000  = 20 if $15,000- $25,000  = 30 if $25,000-$35,000  = 42.5 if $35,000-$50,000  = 62.5 if $50,000-$75,000  = 87.5 if $75,000-$100,000  = 125 if $100,000-$150,000  = 175 if > $150,000 |
| Risk | Self-reported risk tolerance level | + | From 1 to 10 (10 is most likely to take risk) |
| Debt | Self-reported debt level | - | From 1 to 7 (7 is having too much debt by self-assessment) |
| LowTest | A categorical variable for the tested financial literacy level | - | = 1 if answer correctly at most 2/6 questions  = 0 if answer correctly more than 2 questions |
| HighTest | A categorical variable for the tested financial literacy level | + | = 1 if answer correctly 6/6 questions  = 0 if answer fewer than 6 questions correctly |
| Familiar | A proxy discrete variable for the familiarity level with financial mobile apps, created by summing up the numeric answers for questions about familiarity with mobile apps. | + | From -4 to 4 (4 is very frequently use mobile banking, online banking, and financial budgeting apps) |
| LowEd | A categorical variable for the education level | - | = 1 if did not finish high school  = 0 otherwise |
| HighEd | A categorical variable for the education level | + | = 1 if have associate’s, bachelor’s or master degree  = 0 otherwise |
| FinLit | Self-assessed financial literacy level | + | From 1-7 (7 is having very high financial literacy) |

Table 2: Summary statistic of the variables for the dataset and for people who do/do not participate in the stock market

|  |  |  |  |
| --- | --- | --- | --- |
|  | Whole Data Set | Participate | Not Participate |
| |  | | --- | | Variable | | Participation | | Familiar | | White | | Male | | Age | | Income | | Risk | | Debt | | LowTest | | HighTest | | LowEd | | HighEd | | FinLit | | |  |  |  |  | | --- | --- | --- | --- | | Mean | SD | Min | Max | | .380 | .485 | 0 | 1 | | -.456 | 2.022 | -4 | 4 | | .760 | .427 | 0 | 1 | | .455 | .498 | 0 | 1 | | 48.9 | 16.198 | 21 | 70 | | 68.82 | 46.028 | 10 | 175 | | 4.975 | 2.644 | 1 | 10 | | 3.621 | 2.340 | 1 | 7 | | .302 | .459 | 0 | 1 | | .091 | .287 | 0 | 1 | | .015 | .124 | 0 | 1 | | .490 | .500 | 0 | 1 | | 5.23 | 1.268 | 1 | 7 | | |  |  |  |  | | --- | --- | --- | --- | | Mean | SD | Min | Max | | 1 | 1 | 1 | 1 | | -.400 | 2.094 | -4 | 4 | | .0.794 | 0.405 | 0 | 1 | | .557 | 0.497 | 0 | 1 | | 52.0 | 16.00 | 21 | 70 | | 89.73 | 47.523 | 10 | 175 | | 6.1 | 2.354 | 1 | 10 | | 2.91 | 2.247 | 1 | 7 | | 0.193 | 0.395 | 0 | 1 | | 0.161 | 0.368 | 0 | 1 | | 0.004 | 0.064 | 0 | 1 | | 0.616 | 0.486 | 0 | 1 | | 5.683 | 1.056 | 1 | 7 | | |  |  |  |  | | --- | --- | --- | --- | | Mean | SD | Min | Max | | 0 | 0 | 0 | 0 | | -.528 | 1.973 | -4 | 4 | | .0.740 | 0.439 | 0 | 1 | | .393 | 0.488 | 0 | 1 | | 47.0 | 16.026 | 21 | 70 | | 55.99 | 40.004 | 10 | 175 | | 4.285 | 2.574 | 1 | 10 | | 4.057 | 2.288 | 1 | 7 | | .369 | 0.483 | 0 | 1 | | .0.413 | 0.213 | 0 | 1 | | .022 | 0.148 | 0 | 1 | | .413 | 0.492 | 0 | 1 | | 4.953 | 1.306 | 1 | 7 | |

The reason two measurements of financial literacy levels are included is that they do not have a strong positive correlation (0.2747). I hypothesize that people tend to overestimate themselves, as 30.2% of respondents scored low in the test and only 9.1% of respondents scored high in the test, but 25% of respondents claimed that they are at least a 6 on a scale of 7 (table 2). Therefore, it would be interesting to see whether the effects of these two measurements on the dependent variables are different from each other.

As stated above, I expect the sign of the key independent variable, Familiar, to be positive, since people who are more familiar with mobile banking apps are more likely to know about trading applications that would lower investing barriers regarding income or accessibility, so they hypothetically tend to participate in the stock market more. This claim is supported by graph 1 that depicts the distribution of the familiarity level with mobile apps between people who do and people who do not participate in the stock market (Figure 1). To elaborate, the portion of people who are at least at level 2 or above (on a scale from -4 to 4) among people who participate in the stock market is larger than that among people who do not. Besides, the portion of people who are at level -4 or -3 among people who participate in the market is smaller than that among people who do not. Looking at the summary statistics (Table 2), the mean score of familiarity of people who participate is -0.4, greater than -0.53, which is the mean score of people who do not.

Figure 1: Familiarity level with financial mobile apps by people who do and do not participate in the stock market

****

The predicted sign of the White variable is positive, meaning that the hypothesis is that white people tend to participate in the stock market more than non-white people after holding other key factors constant. This hypothesis is supported because there are 79.4% of those who participated in the stock market were white (Table 2). However, I would argue that there is no reason white people statistically significantly tend to participate in the stock market more compared to non-white people after controlling for demographic and financial characteristics. On the other hand, the predicted sign of the Male variable is positive as males tend to be the head of the households and often seek additional sources of income and there are more males in the stock market (55.7% - table 2). However, according to Almenberg and Dreber (2015), the coefficient of the Male variable might not be statistically significant, meaning that males are not significantly more likely to invest in the stock market than females. Nonetheless, as gender and race are key demographic factors, I would still control for them in the model.

Moreover, the expected sign of the Age variable is also positive, as when age increases, people will seek additional sources of income and that people would have greater accumulated capital, so they are more likely to invest in stocks. The expected sign of Income is positive as people are more likely to invest when they have more leisure capital, which will increase with higher income and decrease with higher debt level. The predicted sign of Debt is negative because when people have too much debt, they would prioritize paying off the debt first since the interest rates on their debt are usually higher than the potential return from investing in stocks, so more debts means less likely to invest. The expected sign of Risk is positive because stock investing is inherently risky than other tools of investment, so people who can take more risks would be more likely to participate in stock investing compared to those who usually play it safe. The expected sign of LowTest is negative and HighTest is positive because people who score higher in the financial literacy quiz would have more financial knowledge, and thus are more comfortable and confident in investing, so they are more likely to trade stocks. The same logic applies to why FinLit should have a positive sign. Similarly, I would argue that LowEd has a negative sign and HighEd has a positive sign, as having higher education would allow the respondents to learn advanced economics, finance, and mathematics, thus they will be more likely to trade stocks. On the same page, people with higher education often come from better socio-economic backgrounds compared to those who have lower education, so they are more likely to invest.

The above hypothesis is supported by the fact that there are more people with a medium or low level of education than people with higher education among those who do not participate in the stock market (figure 2). In contrast, there are more people with higher education among people who participate in the stock market. Furthermore, people who have higher education are more confident about their financial knowledge: they self-assessed themselves to have an average score of 5.33 while people who have lower education level only give themselves 4.95 scores on average. All of this points to HighEd having a positive impact on the likelihood of investing, while LowEd having a negative impact.

Figure 2: Education level between people who do and do not participate in the stock market



Not only people who have higher education are more confident about their financial literacy level, they indeed tend to score higher on the test than people with lower education. Among people with an associate’s degree, bachelor’s degree, or higher degree, 13.17% of them answered all questions correctly, compared to only 3.618% of people with the education level of high school/GED degree or less (figure 3). As the interaction between LowTest, HighTest, FinLit and HighEd, LowEd could explain part of the interaction between HighEd, LowEd and the dependent variable, strong linear correlations between these independent variables could bias the model.

Figure 3: Level of education versus the financial literacy score****

When looking at other interactions between independent variables, the risk tolerance level varies between genders, specifically women tend to have lower risk tolerance compared to men. The average risk tolerance level of women is 4.28 while the average risk tolerance level of men is 5.70 on a scale of 10. As there are more males than females among those who participate in the stock market (55.7% versus 44.3%), this can partially explain why the risk tolerance level distributions are different between people who do and do not participate in the stock market. Figure 4 and 5 confirmed this claim, as the distribution of the risk tolerance level between two genders and between people who do and do not participate follow a very similar pattern. As a result, we can hypothesize that women are less likely to participate in the stock market because their risk tolerance level is lower than that of men on average.

Figure 4: Risk tolerance level of people who do and do not participate in the stock market



Figure 5: Risk tolerance level between two genders



Similarly, the self-assessed literacy level also varies between genders (figure 6). Women tend to rate their financial literacy level to be lower than men as their average value is 4.94 while that of men is 5.36 (on a scale of 10). Looking at figure 6, we see that more men think they have a high literacy level than women think about themselves, and a majority of women chose the neutral answer of 5 scores on a scale of 10.

Figure 6: Self-assessed financial literacy level by gender

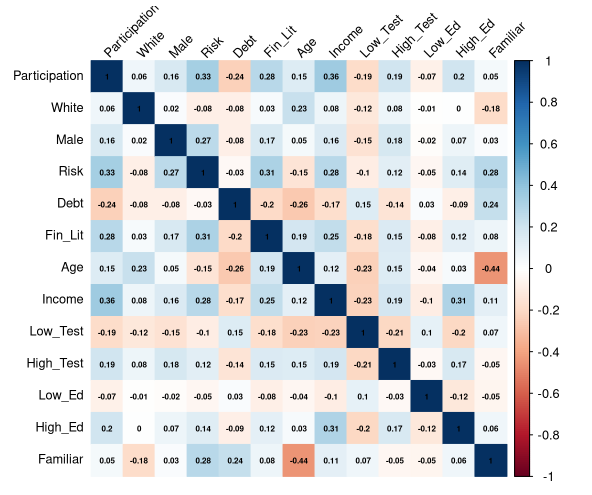
Similar to how the distribution of risk tolerance level between two genders is very much alike the distribution of risk level between people who do and do not participate, the distribution of self-assessed literacy level between genders is also similar to the distribution of that between the two main groups (figure 7). This can be inferred as women are less likely to trade stocks than men because they are less confident in their financial ability. In conclusion, the difference in risk and financial literacy levels between genders account for the difference in risk and financial literacy between people who do and do not participate in the market. This is surprising because the genders of people who do and do not participate are not much skewed to one gender.

Figure 7: Self-assessed financial literacy level by people who do and do not participate in the stock market

****

There might be concerns with strong linear correlations between the independent variables that could bias the model, specifically between gender, risk tolerance, and financial literacy, and between education and financial literacy. However, looking at figure 8, we see that the correlation between Male and Risk is 0.27, between Male and FinLit is 0.17, between HighEd and FinLit is 0.12, between LowEd and FinLit is -0.06, all are weak correlation. Indeed, the strongest correlation is -0.44 between Age and Familiar, which makes sense as older people are less adapted to technology, but this correlation should not be of concern as the two variables can only be considered as moderately correlated.

Figure 8: Correlation matrix between the variables



**Estimation and Results**

In the base Linear Probability Model, all independent variables and the model itself are statistically significant at the conventional levels. The signs of all the coefficients are as predicted. There is no multicollinearity issue with this model as we have believed because all VIF values are under 1.5, much lower than the threshold of 5. However, the adjusted R-squared value is only 0.252, meaning that the independent variables only explain 25.2% of the variation in the likelihood to participate in the stock market. Furthermore, after performing the Breusch-Pagan test, it is confirmed that the model also has a heteroskedasticity issue (BP = 1274.1, df = 12, p-value < 0.01). I have included the White robust standard errors to account for the effect of heteroskedasticity on the statistical significance of the model. However, even though there is no important variable missing, the R-squared value is still lower than the standard threshold of 0.3 for cross-sectional data. One possibility is that the function is trying to estimate the linear relationship between the dependent variable and the independent variables, whereas, with a binary dependent variable, a Logistic Regression Model with a sigmoid function would be a better fit that would allow the same set of independent variables to explain more variation in the dependent variable.

As can be seen from the Base Model (Logit) column, switching to the logistic model helps to increase the R-squared value to 0.338, passing the 0.3 threshold (although there should be some discrepancy between the actual adjusted R-squared value and pseudo R-squared value). Based on the Likelihood Ratio Test, the whole model is also statistically significant at the 99% level. All variables in the base logit model are significant at the 99% level (except for Male, which is significant at the 90% level) and have the correct signs. It can be seen that in contrast with what was expected, gender still has a significant impact on the likelihood to participate in the stock market after controlling for financial literacy level and risk tolerance level. However, if we remove the variable Risk and FinLit (Logit 2), the coefficient of Male becomes much greater by 7 times in magnitude (from 0.057 to 0.371) and the variable becomes significant at the 99% level from the 90% level, while the coefficients and the significance of other independent variables barely change. This means that similar to what has been found by Almenberg and Dreber (2015) and what we have expected after looking at the similar trends in figure 2 and 3, and figure 3 and 4, the gender gap in investing could be mostly explained by the difference in risk tolerance level and financial literacy level between males and females.

From the base logistic model, I was interested in looking at the interaction between Familiar and Age, and between Familiar and Male. The reason behind was that older people are less adapted to technology (this claim is supported by the medium negative correlation of -0.44 between Age and Familiar), so even when holding other factors constant, older people who are more familiar with mobile apps are potentially much more likely to participate in the stock market than older people who are not familiar with them. Similarly, I would also imagine that males who are familiar with mobile apps are more likely to trade stocks than males who are not and females who are. Therefore, from the base model, I wanted to further include Male\*Familiar and Age\*Familiar.

The inclusion of the two new interactive variables leads us to Logit Model 3. The pseudo R-squared value slightly increased from that of Logit Model 2 (0.338 to 0.349), meaning that adding the two new variables is a good idea. The accuracy and sensitivity also increase from Logit Model 2 to Logit Model 3, specifically from 0.740 to 0.748 for accuracy and 0.584 to 0.597 for sensitivity, further indicating that model 3 is better. All variables, including the two new ones, are statistically significant at the 99% level, and the inclusion of the two new variables, interestingly, makes the Male variable significant at the 99% level from the 90% level. Based on the Chi-square test (or the likelihood ratio test), the whole model is also statistically significant at the 99% level. The signs of all variables remain similar to these in model 2. Therefore, Logit Model 3 is my best model for this study.

Table 3: Regression Models (Dependent Variable is Participating in the Stock Market – Binary).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Base Model (LPM) | Base Model (Logit) | Logit 2 | Logit 3 (Best Model) | AME |
| \_cons | -0.228 | -4.080 | -1.775 | -4.024 |  |
|  | (0.0179)\*\*\* | (0.113)\*\*\* | (0.077)\*\*\* | (0.114)\*\*\* |  |
|  | [0.0172]\*\*\* | [0.117]\*\*\* | [0.077]\*\*\* | [0.118]\*\*\* |  |
| Familiar | 0.008 | 0.052 | 0.131 | 0.373 | 0.0079 |
|  | (0.002)\*\*\* | (0.009)\*\*\* | (0.009)\*\*\* | (0.030)\*\*\* | (0.0016)\*\*\* |
|  | [0.0016]\*\*\* | [0.009]\*\*\* | [0.009]\*\*\* | [0.031]\*\*\* |  |
| White | 0.029 | 0.161 | 0.071 | 0.177 | 0.041 |
|  | (0.007)\*\*\* | (0.040)\*\*\* | (0.038)\* | (0.040)\*\*\* | (0.001)\*\*\* |
|  | [0.007]\*\*\* | [0.040]\*\*\* | [0.038]\* | [0.040]\*\*\* |  |
| Male | 0.015 | 0.057 | 0.371 | 0.094 | 0.008 |
|  | (0.014)\*\* | (0.034)\* | (0.031)\*\*\* | (0.034)\*\*\* | (0.006) |
|  | [0.006] | [0.033]\* | [0.031]\*\*\* | [0.034]\*\*\* |  |
| Risk | 0.043 | 0.237 |  | 0.236 | 0.030 |
|  | (1.231)\*\*\* | (0.007)\*\*\* |  | (0.007)\*\*\* | (0.007)\*\*\* |
|  | [0.001]\*\*\* | [0.007]\*\*\* |  | [0.007]\*\*\* |  |
| Debt | -0.030 | -0.166 | -0.177 | -0.167 | -0.029 |
|  | (0.001)\*\*\* | (0.007)\*\*\* | (0.007)\*\*\* | (0.008)\*\*\* | (0.001)\*\*\* |
|  | [0.001]\*\*\* | [0.007]\*\*\* | [0.007]\*\*\* | [0.008]\*\*\* |  |
| FinLit | 0.032 | 0.211 |  | 0.185 | 0.032 |
|  | (0.002)\*\*\* | (0.016)\*\*\* |  | (0.015)\*\*\* | (0.003)\*\*\* |
|  | [0.002]\*\*\* | [0.016]\*\*\* |  | [0.016]\*\*\* |  |
| Age | 0.003 | 0.017 | 0.014 | 0.016 | 0.003 |
|  | (0.0002)\*\*\* | (0.001)\*\*\* | (0.001)\*\*\* | (0.001)\*\*\* | (0.0002)\*\*\* |
|  | [0.0002]\*\*\* | [0.001]\*\*\* | [0.001]\*\*\* | [0.001]\*\*\* |  |
| Income | 0.002 | 0.010 | 0.012 | 0.010 | 0.002 |
|  | (0.000021)\*\*\* | (0.00038)\*\*\* | (0.00037)\*\*\* | (0.00039)\*\*\* | (0.0001)\*\*\* |
|  | [0.000072]\*\*\* | [0.00039]\*\*\* | [0.00037]\*\*\* | [0.00040]\*\*\* |  |
| Low\_Test | -0.034 | -0.239 | -0.278 | -0.275 | -0.047 |
|  | (0.007)\*\*\* | (0.039)\*\*\* | (0.037)\*\*\* | (0.395)\*\*\* | (0.007)\*\*\* |
|  | [0.006]\*\*\* | [0.039]\*\*\* | [0.037]\*\*\* | [0.040]\*\*\* |  |
| High\_Test | 0.101 | 0.457 | 0.605 | 0.474 | 0.081 |
|  | (0.010)\*\*\* | (0.058)\*\*\* | (0.056)\*\*\* | (0.058)\*\*\* | (0.010)\*\*\* |
|  | [0.012]\*\*\* | [0.058]\*\*\* | [0.056]\*\*\* | [0.034]\*\*\* |  |
| Low\_Ed | -0.051 | -0.728 | -0.764 | -0.678 | -0.116 |
|  | (0.023)\*\* | (0.193)\*\*\* | (0.186)\*\*\* | (0.195)\*\*\* | (0.033)\*\*\* |
|  | [0.016]\*\*\* | [0.181]\*\*\* | [0.180]\*\*\* | [0.183]\*\*\* |  |
| High\_Ed | 0.058 | 0.317 | 0.333 | 0.332 | 0.057 |
|  | (0.006)\*\*\* | (0.034)\*\*\* | (0.032)\*\*\* | (0.034)\*\*\* | (0.006)\*\*\* |
|  | [0.006]\*\*\* | [0.034]\*\*\* | [0.033]\*\*\* | [0.034]\*\*\* |  |
| Familiar\*Age |  |  |  | -0.008 |  |
|  |  |  |  | (0.0005)\*\*\* |  |
|  |  |  |  | [0.00060]\*\*\* |  |
| Familiar\*Male |  |  |  | 0.115 |  |
|  |  |  |  | (0.016)\*\*\* |  |
|  |  |  |  | [0.016]\*\*\* |  |
| N | 22,368 | 22,368 | 22,368 | 22,368 | 22,368 |
| Adj. R-squared | 0.2525 |  |  |  |  |
| Pseudo R-squared |  | 0.338 | 0.258 | 0.349 | 0.349 |
| SEE | 3937.243 |  |  |  |  |
| F-ratio | 630.7\*\*\* |  |  |  |  |
| SSR | 1332.936 |  |  |  |  |
| Log likelihood | -12310.52 |  |  |  |  |
| Model LRT |  | 6384.08\*\*\* | 4697.48\*\*\* | 6636.84\*\*\* | 6636.84\*\*\* |
| Accuracy  (Sensitivity) |  | 0.7402  (0.5845) | 0.7124  (0.4905) | 0.7484  (0.5971) | 0.7484  (0.5971) |

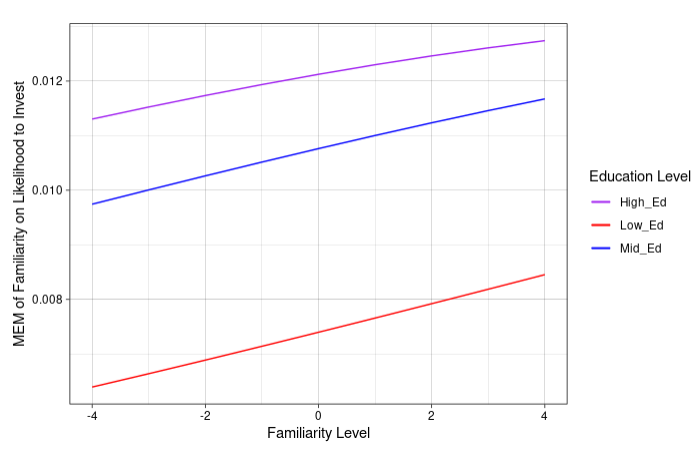
(Standard errors in round bracket, White robust standard errors in square bracket. p<0.01: \*\*\*, p<0.05: \*\*, p<0.10: \*)

It can be seen in table 3 that the sensitivity of 59.71% of model 3 is low, while the accuracy of 74.84% is moderate. In fact, the specificity of model 3 is pretty high at 84.11%, and since there is a trade-off between sensitivity and specificity, I decided to lower the specificity to make the sensitivity of the model higher by changing the decision threshold level. The decision threshold for the base logit model, model 2, and model 3 was at 0.5 by convention. After trials, I decided that the threshold of 0.4 is the best, which would give the accuracy of 73.66%, sensitivity of 71.63%, and specificity of 74.91%, all are moderately high. If we further lower the threshold, the accuracy and specificity would be much lower than the sensitivity and the model would lose the balance in its predictive power.

The positive sign and the significance level of the Familiar variable confirmed my hypothesis that familiarity with mobile financial services inspired people to participate in stock trading. To elaborate, by looking at the AME, one unit increase in the familiarity with financial mobile apps score would increase the probability of participating in the stock market by 0.79 percentage point. In other words, if the same person who is moderately familiar with financial mobile app (Familiar = 0) became extremely familiar with the apps (Familiar = 4), the likelihood of the person participating in the stock market only increases by 3.16 percentage point, much smaller than the effect of other independent variables. Although the economic significance of the familiarity with mobile apps is small, it should be taken into account that mobile apps are much easier to learn compared to increasing a persons’ education level, or financial knowledge.

Regarding other variables, the factors that have the greatest effect on the likelihood of investing in the stock market in decreasing order are the education level of the participants, their scores on the financial literacy test, their race, their self-assessed financial literacy level, their risk tolerance level, then their debt level. Most significantly, the likelihood of participating in the stock market for a person who did not finish high school is 11.6 percentage point lower compared to a person who has a GED or high school degree, and a person who has an associate’s, bachelor’s, or master/Ph.D. degree would have 5.7 percentage point higher in the likelihood to participate in the stock market compared to people who have a high school or GED degree. When keeping other factors constant at means, one unit increase in the familiarity level with financial mobile apps would always increase the likelihood to participate in the stock market for people of all levels of education (figure 9).

Figure 9: MEM of Familiar on the Likelihood to Invest in different level of education



A person who answered correctly 6/6 questions on the financial literacy test would have 8.1 percentage points higher in the likelihood to trade stocks, while a person who answered correctly 1 or 2 questions on the test would have 4.7 percentage points lower in the likelihood to trade stocks, all compared with a person who answered from 3 to 5 questions correctly. Furthermore, surprisingly, after controlling for demographic factors such as age, income, and educational level, a white person would have a 4.1 percentage points higher in the likelihood to invest in stocks compared to non-white people. This racial gap in investing was hypothesized to be attributed to differences in saving behavior and investment choice of the minorities in previous literature (Choudhury, 2002) as non-white households tend to invest in less risky assets. However, this should not be the case here as Logit Model 3 controls for risk tolerance level as well. Consequently, the investing gap between races remained unexplained.

The next factor that has the highest influence on the dependent variable is FinLit. To elaborate, an increase in the score of FinLit (on a scale from 1 to 7) would increase the likelihood to invest by 3.2 percentage points on average. It can be seen that the effect of the respondents’ self-assessed financial literacy level and their score on the financial literacy quiz are very different, particularly the scores on the test have more predictive power or are more reliable than their self-assessed literacy level. This further supported my aforementioned claim that respondents tend to overestimate their financial literacy. Nonetheless, both of them are still among the factors that have the most influence on the likelihood to participate in the stock market.

Next, one unit increase in the risk tolerance level (on a scale from 1 to 10) of the respondents would lead to an increase of 3.0 percentage point in the likelihood to invest in stocks, and one unit increase in the self-assessed debt level (on a scale from 1 to 7). It is worth noting that debt and income both influence the wealth of the respondents, and thus influence the likelihood to invest in stocks of the respondents, but when we look at the AME, the debt level has a much stronger effect compared to the income level. However, the unit of the Income variable is in thousands of dollars, and an increase of one thousand dollars is much easier than a one-unit increase in the debt level on a scale from 1 to 7. To put it comparably, a $10,000 increase in Income would increase the likelihood of investing by 2.0 percentage points in the probability of participating in the stock market, approximately the effect of Debt.

Furthermore, one increase in the familiarity level with financial mobile apps decrease the probability by 0.17 percentage point for a woman and increase the probability by 1.84 percentage point for a man on average. It is interesting to see the opposite effect of familiarity on females and males, however, the average marginal effect for females are not significant at conventional levels (SE = 0.0021, z = -0.7932, p-value = 0.4276), while that of males are statistically significant at the 99% level (SE – 0.0022, z = 8.4399, p-value < 0.01). Therefore, we could say that the average marginal effect of one increase in the familiarity level on a scale of -4 to 4 does not have a significant effect on the likelihood of a female to invest.

Finally, when keeping other factors constant at their means, one unit increase in the familiarity level with financial mobile apps for a 48.9-year-old person (the mean age) would increase the likelihood to invest by 1.13 percentage points. However, this effect is different across age groups. To be more specific, when keeping other factors constant at means, the effect of one unit increase in the familiarity level would have a higher effect on the likelihood to participate in the stock market for a 21 years old (4.44 percentage point), compared to older people. For example, at 40 years old it would be 2.41 percentage point, and at 70 years old it would decrease the likelihood by 2.56 percentage point. Similarly, the average marginal effect of one unit increase in familiarity on the dependent variable also decreases when age increases (at 21: an increase of 4.02 percentage point, at 48.9: an increase of 0.98 percentage point, at 70: a decrease of 1.93 percentage point). All of the above effects are statistically significant at the 99% level, which means that the effects are unlikely to be random. This is surprising because I would hypothesis that for older people, keeping other factors constant, a person who is more familiar with mobile apps should be more likely to invest in stocks, not less likely. Currently, I have no alternative hypothesis that could explain this phenomenon.

To sum up, it is interesting to see that the factors that have the highest influence on the likelihood to invest in stocks are factors regarding the financial literacy level or educational level of the respondents. On the other hand, demographic factors such as age, income, debt, or gender have less effect on the dependent variable, except race. The average marginal effect of my key independent variable has small economic significance (0.79 percentage point), but the effect of it varies a lot with different groups of respondents. For example, as stated above, its marginal effect on a 21-year-old person when keeping other factors at their average is high (4.4 percentage point), around the difference in likelihood to invest between white and non-white people.

To further elaborate on this point, I would make several predictions about hypothetical people. I would arrive at the possibility to invest in stock by applying the function

**Where Z is the value obtained after plugging in the values for independent variables using Logit Model 3.**

For example, for myself as a female who is non-white, very familiar with financial mobile apps (Familiar = 4), with a risk tolerance of 5 on a scale of 7, debt level of 1 on a scale of 10, self-assessed financial literacy of 7, age of 21, lowest bracket of income (Income = 10), has a high score on the financial literacy quiz (answered correctly 6/6 questions) and medium education (as I have not yet obtained my bachelor’s degree), Z would be 3.614, so the probability would be 97.38%, which is very high. Given the decision threshold of 40%, the model would predict I do invest in the stock market, which is indeed true. On a different scenario, for a male who is non-white, familiar with financial mobile apps (Familiar = 3), with the risk tolerance of 2 on a scale of 7, the debt level of 1 on a scale of 10, self-assessed financial literacy of 2, age of 23, bracket of income from $50,000-$75,000 (Income = 62.5), high education (a bachelor’s degree), has the medium score on the financial literacy quiz (answered 5/6 correctly), Z would be 1.098, the probability would be 74.99%. This prediction is false for my friend who does not invest in the stock market given the decision threshold of 40%, but it is a good thing that this probability is lower than what the model gave for my previous scenarios. As the accuracy and sensitivity of model 3 are around 70%, false predictions are expected. Finally, for a hypothetical female who is white, very familiar with financial mobile app (Familiar = 4), with a risk tolerance of 2 on a scale of 7, self-assessed financial literacy of 4 on a scale of 10, age of 35 years old, bracket income from $50,000-$75,000 with medium test score and medium education, Z would be -3.04 and the probability would be very low (4.56%) and the model would predict that she does not invest in the stock market.

**Conclusion and Implications**

To sum up, similar to previous research (Bogan, 2008), the result of this study confirmed the hypothesis that familiarity with financial mobile apps would increase the likelihood to invest in the stock market, specifically by 0.79 percentage point for every one-unit increase in the familiarity score on a scale from -4 to 4 on average. This factor does not have great economic significance as the magnitude of the influence is small, nonetheless, it should be taken into account that mobile apps are much easier to learn compared to increasing a persons’ education level, or financial knowledge. This could be used as another tool to encourage investing targeting specific groups of people, especially young people because familiarity has the strongest effect on them. This could also mean that in the future when there are more financial mobile apps, specifically more trading apps like Robinhood, there should be more young investors joining the stock market. This hypothesis could be confirmed by future research that uses two different sets of data from two different periods, one with fewer trading apps available and one with more trading apps. In fact, there are datasets that were from the State-to-State survey by the National Financial Capability Study in 2013 and 2015 (the dataset that was used in this research is from 2018). As Robinhood, the pioneer of no fee, no minimum balance brokerage app was founded in 2013, future research could potentially compare the proportion of investors among the respondents between 2013 and 2018, and how the key factors in this research affect the likelihood to invest in stocks differently in 2013.

There are several other interesting implications from this research. First, similar to previous research (Almenberg and Dreber, 2015), controlling for financial literacy and risk tolerance level lessens the effect of gender on likelihood to invest; but unlike the conclusion from previous research, there is still a significant gender gap in investing (although it is small). Secondly, there is also a significant gap in the likelihood to invest between white and non-white people of 4.1 percentage points even after holding other factors constant that remains unexplained. Thirdly, the strongest factors that would (positively) influence the likelihood to trade stocks are related to knowledge, specifically, financial literacy level and education level, while other demographic factors do not matter as much. Thirdly, old people who are more familiar with financial mobile apps are less likely to invest in stocks than old people who are less familiar with the apps, the phenomenon that remains unexplained as well. And finally, there is a weak correlation between people’s self-reported financial literacy level and their actual financial literacy level that is calculated from the number of correct answers on the given quiz, and I hypothesized that people tend to overestimate their financial literacy level.

I would argue that the best model in this research is reliable given the pseudo R-squared value, the significance level of the variables and the model itself, and the accuracy/sensitivity of the model. There should be no missing important variable, however, there are other variables that could be added to the model to improve its predictive power regarding accuracy/sensitivity. For example, I wish to include a variable that indicates the leisure time a person has, as investing in stock requires a lot of time to keep track of the news, check the stock market, and read additional materials to find out about potential stocks or learn about trading strategies. The occupation of a person would possibly affect the likelihood to invest in stocks as well, as a person who works in banking would be more likely to invest in stock than a pilot, for example. Other than that, the model could also have higher predictive power if it were trained on a better dataset. In this dataset, variables such as age and income that could be numerical are given in the categorical form (age and income bracket), which does not contain as much data as when they are specific numerical values. Furthermore, the survey that obtained this dataset was distributed online, which could potentially bias the model because my key independent variable relates to mobile and online applications. Nonetheless, as I have mentioned, the model is still reliable without omitted variable bias, outliers, and multicollinearity problems.

**References:**

Almenberg, J. and Dreber, A. (2015). “Gender, stock market participation and financial literacy.” *Economic Letters*, 137, 140-142. <https://www.sciencedirect.com/science/article/pii/S0165176515004115>

Bogan, V. (2008). “Stock Participation and the Internet.” *Journal of Financial and Quantitative Analysis*, 43(1), 191-212. <https://www.jstor.org/stable/27647344?seq=1#metadata_info_tab_contents>

Choudhury, S. (2002). “Racial and ethnic differences in wealth holdings and portfolio choices.” *ORES Working Paper*, 95.

<https://www.ssa.gov/policy/docs/workingpapers/wp95.html>

Fitzgerald, M. (2020). *Many Americans used part of their coronavirus stimulus check to trade stocks*. CNBC. Retrievied from: <https://www.cnbc.com/2020/05/21/many-americans-used-part-of-their-coronavirus-stimulus-check-to-trade-stocks.html>

Huang, D. (2015). *Young, poor and looking to invest? Robinhood is the app for that.* The Wall Street Journal. Retrieved from <https://blogs.wsj.com/moneybeat/2015/01/06/young-poor-and-looking-to-invest-robinhood-is-the-app-for-that/>

Robinhood. Retrieved from <https://robinhood.com/us/en/>

Rooij, M., Lusardi, A., and Alessie, R. (2011). “Financial Literacy and Stock Market Participation.” *Journal of Financial Economics*, 101(2), 449-472. <https://www.sciencedirect.com/science/article/pii/S0304405X11000717>