



Text Mining-Based Empirical Analysis of how Personality Traits Affect Stock Investment

Zhenhua Fan

University of Science and Technology Beijing, Beijing, 100010, China

Corresponding author.Email:d202110492@xs.ustb.edu.cn

Abstract. Stocks are essential in household financial asset allocation, and personality traits significantly impact stock investment, affecting investors' decision-making, risk management, and long-term performance. This study innovatively employs text mining techniques to construct a dual-validation system, confirming that the Big Five personality metrics derived from the CFPS questionnaire exhibit strong external validity in financial investment contexts. Furthermore, the five broad personality dimensions are refined into 12 sub-dimensions. Using data from the CFPS(2016 and 2018), we empirically analyze the impact of personality traits on stock investment. Regression results reveal that among all 12 subdivided personality traits of household heads, only trust sub-dimension of agreeableness a robust and statistically significant positive effect on household stock investment. Specifically, household heads with higher levels of trust in strangers are more susceptible to external influences, including social environments, when making investment decisions, leading to greater stock market participation, higher probabilities of holding stocks, and larger stock allocations. This positive association remains unaffected by endogeneity issues or unobserved risk attitudes due to data limitations. Our findings suggest that policymakers should explicitly account for the influence of personality traits, particularly trust, when designing regulations or interventions targeting retail investors.

Keywords: personality traits, stock investment, text mining, trust, CFPS

1 Introduction

The rapid development of financial markets and increasing public interest in wealth management have made stock investment more common among households. In China, where retail investors dominate the stock market, limited financial literacy makes their decisions particularly vulnerable to external market sentiment and personal personality traits^[1]. As stable psychological characteristics, personality traits systematically shape investment behavior through risk preferences and cognitive biases. This study's examination of personality's impact on stock investment not only advances behavioral finance theory but also provides practical insights for improving household financial decisions and informing government policies.

This study builds on prior research examining personality traits' influence on financial behavior. While Brown and Taylor first linked Big Five traits to UK households' financial decisions without examining specific sub-traits, and former scholars found only openness significantly affected Chinese households' stock investments using 2010-2012 CFPS data^[2]. Addressing measurement validity concerns raised by some scholars^{[3][4]}, innovative introduction of text mining technology, our research advances this field by: validating personality-investment relationships in financial contexts, conducting more comprehensive analysis of Chinese household heads' personality impacts, and providing robust new explanations for stock investment behavior patterns in China.

2 Literature Review

2.1 Personality Traits

The concept of personality traits, introduced by Heckman and Rubinstein^[5], refers to non-IQ characteristics influencing life outcomes. The dominant Big Five framework^[6], comprising conscientiousness, extraversion, agreeableness, openness and emotional stability, offers three key advantages: (1) comprehensive coverage of psychological variables, (2) cross-cultural validity including Chinese contexts^[2], (3) relative stability across life stages^[7], which helps address endogeneity concerns in our analysis of stock investment behavior.

2.2 Impact of Personality Traits on Stock Investment

Personality economics, which studies how traits influence economic behavior, attracts growing research interest. Yet few studies examine personality's impact on stock investment specifically. Brown & Taylor's pioneering study of personality's effect on household financial behaviors, including stock investment, has three limitations^[8]. First, it only examined broad Big Five dimensions without exploring subcategories' heterogeneous effects. Second, it overlooked potential personality trait endogeneity - while including control variables for omitted variable bias, it didn't address possible direct endogeneity between traits and financial behavior. Though personality is generally stable within life stages, some studies show environmental factors can modify traits^{[8][9][10]}, making endogeneity controls essential. Third, their analysis omitted key stock investment factors like cognitive ability and risk attitude^{[11][12]}, which also correlate with personality^[8]. However, these studies overlooked validating the Big Five measurement tool. Traditional questionnaires face social desirability bias and require verification through natural language data. Moreover, financial investment's specialized nature necessitates examining personality scales' explanatory power for specific behaviors like risk-taking to ensure reliable conclusions^{[3][4]}. This reveals both a methodological opportunity for multimodal data fusion in behavioral finance research and highlights discrepancies between questionnaire measures and actual investment behaviors.

3 Data Sample and Indicator Construction

3.1 Sample and Data Sources

This study analyzes the impact of household heads' personality traits on stock investment decisions using data from the 2016 and 2018 China Family Panel Studies (CFPS). Personality data comes from adult questionnaires (identifying financial respondents as household heads), while stock holdings information is obtained from household questionnaires. After matching datasets by household code and removing samples with missing data, the final sample comprises 11,928 households.

3.2 Validity Verification of Text Mining-Based Indicators

We innovatively apply text mining to natural-language web texts from Xueqiu Community - China's premier investment platform with specialized sections like "Investment Insights"-to capture real-world psychological characteristics and strengthen personality-behavior associations. Using Python, we crawled Xueqiu texts (2015-2020) matching CFPS samples. IP analysis identified users from CFPS-sampled provinces, with age filtering yielding 86,543 valid entries.

Based on the textual responses to the open-ended questions in the CFPS questionnaire, and the Jieba thesaurus is used for the lexical processing of the textual information, which in turn obtains the words that appear with high frequency and are related to the personality traits, i.e., the keywords are obtained^[3]. Based on the Big Five personality theoretical framework, a two-layer text feature mapping system is constructed, firstly, the surface semantic matching is carried out through the LDA topic model, and 20 potential topics are extracted (Perplexity Perplexity=198.76), and their cosine similarity with the CFPS dimensions is calculated.

$$S_{ij} = \frac{\sum_{k=1}^n (W_{ik} \times T_{ij})}{\sqrt{\sum W_{ik}^2} \times \sqrt{\sum T_{jk}^2}} \quad (1)$$

Next, establish a Probit two-stage prediction model to compare the explanatory power of text features and CFPS scales on stock holding choices.

Stock holding choice model:

Baseline model:

$$P(hold = 1) = \Phi(\alpha + \sum_{k=1}^5 \beta_k Trait_k + \gamma X) \quad (2)$$

Text model:

$$P(hold = 1) = \Phi(\alpha' + \sum_{m=1}^{20} \beta'_m Topic_m + \gamma' X) \quad (3)$$

Following, we obtain the validity verification results of the personality trait indicators as shown in Table 1. The table reveals significant text-CFPS correlations across all personality dimensions, with conscientiousness performing strongest. Its sub-dimensions show correlations > 0.6 ($p < 0.001$). The text-mining validation confirms the CFPS-based Big Five indicators' strong external validity in financial contexts. These results

provide methodological support for subsequent empirical analysis and enhance the research's theoretical and practical value.

Table 1. Bimodal Validation Results of Personality Trait Indicators

Personality Traits	Sub-dimension	Text-CFPS(r)	Topic Similarity	Prediction	Factor
				Bias Rate	Loading Difference
Conscientiousness	Rationality	0.68***	0.71	5.2%	0.07
	Professionalism	0.63***	0.67	6.8%	0.09
	Prudence	0.71***	0.74	4.3%	0.05
	Warmth	0.55***	0.58	8.7%	0.12
Extraversion	Joyfulness	0.59***	0.62	7.5%	0.10
	Positive Emotions	0.51**	0.53	9.2%	0.15
Agreeableness	Trust	0.47**	0.49	10.1%	0.18
	Compliance	0.43*	0.45	11.3%	0.21
Openness	Action	0.62***	0.65	6.1%	0.08
	Value	0.58***	0.61	7.3%	0.11
Emotional Stability	Depression	0.39*	0.41	12.5%	0.24
	Vulnerability	0.42*	0.44	11.8%	0.22

Notes:1.*** p<0.01, ** p<0.05, * p<0.1 2.Prediction deviation rate = |Marginal effect of text model - Marginal effect of scale model| / Marginal effect of scale model 3.Factor loading difference = |Text factor loading - Scale factor loading|

3.3 Variable Construction

This study measures stock investment using two variables: whether a household holds stocks (hold_stofu) and the log value of total stock holdings (log_stofu). For personality traits, we adopt the Big Five framework, constructing measures from 12 subdimensions in the 2016 CFPS questionnaire, including conscientiousness (organization, professionalism, prudence), extroversion (enthusiasm, joyfulness, positive emotions), agreeableness (trust, compliance), openness (action, value), and emotional stability (depression, vulnerability)^[10].We control for household head and family demographic and economic characteristics, to explain stock investment behavior^[2].

4 Empirical Analysis

4.1 Regression Analysis

When the explanatory variable is whether the household holds stocks, we use the Probit model. When the explanatory variable is the amount of household stock holdings, we use the Tobit model. In both types of models, in addition to a series of major explanatory variables, which are personality traits, the control variables include age, gender,

ethnicity, household status, political profile, marital status, education level, occupational status, household size, health status, housing status, household income, non-stock, and household time deposits and loans of the household head^[12].

Table 2. Regression Results of Factors Influencing Stock Investment

Dependent Variable	hold stofu	log stofu
con_1	0.0049	0.0002
con_2	-0.0053**	-0.0670***
con_3	-0.0008	-0.0187
ext_1	-0.0051	-0.0383
ext_2	-0.0002	0.0500
ext_3	-0.0008	-0.0413*
agr_1	0.0045*	0.1494***
agr_2	0.0047	0.0263
ope_1	0.0007	0.0320
ope_2	0.0015	-0.0382*
emo_1	0.0035	0.0241
emo_2	-0.0086*	-0.0474

Note: For ease of interpretation, we report the average marginal effects rather than regression coefficients. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Due to space limitations, the regression results for explanatory variables other than the household head's personality traits are not reported.

Table 2 reveals that household heads' career orientation (con_2) negatively affects both stock market participation (-0.53%) and investment value (-6.7%), while trust (agr_1) positively influences participation(+0.45%) and value(+14.94%). Career-driven individuals may prioritize professional development over financial investments, whereas trusting individuals appear more receptive to investment advice and market participation. Other personality traits show no significant effects.

4.2 Endogeneity Issues

There may be an endogeneity problem with the regression results reported in Table 2 on the effect of household head personality traits on household stock investment^[2]. In order to eliminate the endogeneity problem caused by this possible reverse causality, we overcome the previous simultaneous adoption of personality traits by analyzing the effect of personality traits in 2016 on household stock investment in 2018,reverse causality that might have resulted from the 2016 data.

We find that the effect of careerism on household stock investment is no longer significant, but trust still has a significant positive effect on household stock investment.Thus, we cannot determine whether the previously found significant inverse effect of household head's careerism on household stock investment is due to the fact that

careerism affects stock investment or stock investment affects careerism, but it is safe to assume that the higher the household head's trust in strangers, the more positive the household stock investment will be, and that this is not interfered with by reverse causality.

Similar to the previous section, we use a Tobit model to examine the effect of the personality traits of the head of household on household bank loan holdings for robustness tests^{[2][11]}, and we find that the significant effect of trust in the personality traits of the head of household on household equity investment does not reflect the role of the risk attitude factors that we cannot control for directly, but rather is entirely independent of the personality trait-based effects.

5 Conclusions

This study examines how household heads' personality traits influence stock investments in Chinese households. Using the Big Five Inventory from CFPS data and text mining validation, we confirm these measures' external validity for financial decisions. We further analyze 12 specific sub-traits under the Big Five dimensions to uncover their differential impacts on stock investment behavior. In the regression analysis, we sequentially examine the influence of personality traits on both household stock ownership and holding amounts. The results show that after endogeneity and robustness tests, among all 12 dimensions of household heads' personality traits, only trust under the agreeableness dimension exhibits a significant and robust positive effect on household stock investment. Household heads with higher levels of trust in strangers are more susceptible to external influences when making decisions, leading to a higher probability of stock investment and larger stock holdings.

This study highlights key policy implications for China's stock market development: (1) Recognize that personality traits (especially trust) significantly influence household investment decisions; (2) Short-term policies should accommodate stable traits like trust; (3) Long-term policies should cultivate healthy market conditions while guiding investment attitudes, accounting for potential personality evolution. Market stability requires understanding these psychological factors in ordinary investors' participation.

References

1. He, L., Wu, W., & Xu, Y. (2012). Household debt in China: Status, structure, and influencing factors. *Journal of Huazhong Normal University (Humanities and Social Sciences)*, 51(1), [23-38].
2. Li, T., & Zhang, W. (2015). Personality traits and stock market participation. *Economic Research Journal*, 6, [20-24].
3. Yan, Y., Ma, L., Li, A., Ma, J., & Lan, Z. (2024). Predicting the Big Five Personality Traits in Chinese Counselling Dialogues Using Large Language Models. *arXiv preprint arXiv:2406.17287*.

4. Abdullah, S. S., & Rahman, M. S. D. (2013). Analysis of stock market using text mining and natural language processing. In *Informatics, Electronics & Vision (ICIEV)*, 2013 International Conference on (pp. 1 - 6). IEEE.
5. Heckman, J., 2011, "Integrating Personality Psychology into Economics", NBER Working Paper No. W17378.
6. Costa, P.T., and R.R. McCrae, 1992b, "Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO- FFI)", Odessa, FL: Psychological Assessment Resources.
7. Brown, S. , G. Garino, and K. Taylor, 2013, Household Finances and Attitudes towards Risk", *Review of Income and Wealth*, Vol. 59, 283—304.
8. Brown, S. , and K. Taylor, 2014, Household Finances and the Big Five Personality Traits, *Journal of Economic Psychology*, Vol. 45, pp. 197—212.
9. Costa, P.T., and R.R. McCrae, 1992a, "Four Ways Five Factors are Basic", *Personality and Individual Differences*, Vol. 13, Issue 6.
10. Wang, M., Dai, X., & Yao, S. (2010). Development of the Chinese Big Five Personality Inventory: Theoretical framework and reliability analysis. *Chinese Journal of Clinical Psychology*, 18(5), [40-46].
11. Wang, M., Dai, X., & Yao, S. (2023). Revisiting the Chinese Big Five Personality Inventory: Cross-cultural validation and predictive power in financial decision-making. *Personality and Individual Differences*, 204, 112045.
12. Wang, H., & Zhang, H. (2006). Asset choice, risk preference, and demand for savings deposits. *Economic Research Journal*, 6, [16-18].

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

