

Article

Decomposing the Household Herding Behavior in Stock Investment: The Case of China

Yung-Ching Tseng ¹, I.-Fan Hsiao ^{2,*} and Guo-Chen Wang ³¹ Department of Civic Education and Leadership, National Taiwan Normal University, Taipei 106, Taiwan² Department of Banking and Finance, Tamkang University, New Taipei City 251, Taiwan³ Chung-Hua Institution for Economic Research, Taipei 10672, Taiwan; gcwang@cier.edu.tw

* Correspondence: 153898@o365.tku.edu.tw

Abstract: Financial studies on the herding effect have been very popular for decades, as detecting herding behavior helps to explain price deviations and market inefficiencies. However, studying the herding effect as a single influencing factor is believed to be insufficient to explain the changes in investment behavior, as the herding effect itself may be caused by other influencing factors. In other words, the issue must be studied alongside other factors. In this study, we adopt the quantile regression model to comprehensively understand the herding effect's influence on household investment in China, and the empirical results indicate that herding behavior leads to different investment outcomes for households in different scenarios. In this analysis, we consider a variety of household characteristics, such as income level and risk tolerance, to provide a nuanced understanding of investment behavior. Additionally, in this study, we explore the interaction between herding behavior and macroeconomic variables. Nevertheless, the results suggest that, if herding behavior can be reduced by the head of the household, profitability can be increased, or at the very least, losses can be reduced.

Keywords: the herding effect; household investment; quantile regression; quantile regression-based decomposition

JEL Classification: D14; D91; G41; G53



Academic Editor: Julie Le Gallo

Received: 7 March 2025

Revised: 24 April 2025

Accepted: 6 May 2025

Published: 12 May 2025

Citation: Tseng, Y.-C., Hsiao, I.-F., & Wang, G.-C. (2025). Decomposing the Household Herding Behavior in Stock Investment: The Case of China. *Econometrics*, 13(2), 21. <https://doi.org/10.3390/econometrics13020021>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Financial studies on the herding effect have been popular for decades, as detecting herding behavior helps to explain price deviations and market inefficiencies. Herding behavior occurs when investors ignore the information they have obtained, and instead of making appropriate decisions for their own good, they follow the actions of others (Christie & Huang, 1995). When individuals fail to meet their expectations of investment, it is considered more sensible to replicate others' practices and expect to ultimately achieve a better performance (Demirer & Kutan, 2006). Therefore, the existence of herding effects has jeopardized the market efficiency assumption, as it requires that all information be shared by all investors and rational decisions be made. In fact, individuals, especially non-specialists, do not invest rationally based on the obtained information; instead, they simply follow the actions of others. While the efficient market hypothesis assumes that investors are rational and markets are efficient (Fama & French, 1992), the presence of herding behavior challenges its legitimacy. In fact, herding behavior is one of the major drivers pushing stock prices away from their fundamental values, namely, undervalued and overvalued trades. The mispricing of stocks poses significant risks to the stability and

efficiency of the stock market. Several causes of investors' herding behavior, which affects stock markets directly and affects economic growth indirectly, have been identified (Tan et al., 2008). Herding behavior may cause market failure through different kinds of market disorders, such as significant information loss, insufficient information gathering, reduction in decision-making abilities, higher volatility in asset prices, mispricing, and short-term bubbles in the stock market. In fact, the herding effect may even challenge fundamental financial theory.

In recent decades, the Chinese stock market has experienced significant growth, marked by key milestones such as the establishment of the Shanghai and Shenzhen Stock Exchanges in the early 1990s (Pavlidis & Vasilopoulos, 2020). Economic factors, including rising incomes, urbanization, and improved financial literacy, have further driven household investments, while sociocultural influences, such as the cultural emphasis on savings and the impact of social networks, have also been significant (Liu & Zhang, 2021; X. Wu & Zhao, 2019). Additionally, household background variables, such as gender, educational level, age, and asset status, are influential factors worthy of exploration and have been the subject of extensive scholarly research (Lin, 2011; Talpsepp & Tänav, 2021; W. Li et al., 2020; L. H. Li et al., 2021). Despite these positive trends, challenges such as market volatility and limited financial knowledge persist. However, opportunities presented by technological advancements in fintech and mobile trading apps continue to facilitate stock market participation (Wang et al., 2022; Zhang et al., 2021). Understanding these dynamics is essential to informing policies and strategies to promote stable and informed investment practices among Chinese households.

The first model that was developed to examine the herding effect is known as the cross-sectional standard deviation (CSSD) method (Christie & Huang, 1995). The model is based on the concept of herding behavior, where investors prefer to follow the financial decisions of others, rather than basing their actions on their own choices. However, if low-dispersion statistics indicate the presence of the herding effect while high-dispersion statistics do not, then the herding effect cannot be easily captured. Subsequently, corrections and adjustments have been made by researchers to more accurately estimate the herding effect based on objective data. However, studying an influencing factor of the herding effect in isolation is considered insufficient to explain changes in investment behavior. As the herding effect itself may be caused by other influencing factors, the issue must be studied simultaneously with other factors.

As advanced research has shown, understanding the herding effect in the market can be beneficial for future investment (Saastamoinen, 2008). Therefore, the goal of this study, instead of simply verifying the existence of the herding effect, as in most existing studies, is to examine how herding behavior influences investment profitability beforehand, as investors have actively admitted that they have engaged in herding behavior.

However, most empirical studies on herding behavior have relied on objective market-level indicators, such as return dispersion or beta herding, which do not reflect the individual-level subjective inclination to follow others. Such limitations leave a critical gap in understanding the psychological motivations behind herding.

To address this, our study introduces a novel self-reported measure of herding behavior based on individual responses from the China Household Finance Survey (CHFS), allowing us to directly observe how investors perceive their susceptibility to peer influence. This subjective metric offers a complementary and deeper perspective compared to conventional proxies and enhances our ability to analyze herding behavior within diverse demographic and socioeconomic contexts.

Given China's rapidly growing and retail investor-dominated stock market, understanding household-level herding through a behavioral lens is both timely and important for informing investor education and policy intervention strategies.

While previous literature extensively documents herding behavior using objective proxies such as cross-sectional dispersion of returns or beta herding (Christie & Huang, 1995; Hwang & Salmon, 2006), these methodologies predominantly capture aggregate market phenomena rather than individual psychological tendencies. As a result, they provide limited insights into the personal perceptions and decisions underlying herding behavior.

This study distinguishes itself by incorporating a novel self-reported herding measure from survey responses, directly capturing investors' subjective inclination to follow peers. Such subjective data, rarely adopted in existing studies, provide richer psychological insights and capture nuances that market-based measures may overlook.

Furthermore, we adopt the quantile regression approach, allowing the analysis of how herding behavior's influence varies across different segments of the investor distribution. Traditional analyses using ordinary least squares (OLSs) regression typically assess only the average effect, ignoring heterogeneous effects among investors with differing investment outcomes. By applying quantile regression, we can explore nuanced impacts at various investment performance quantiles, significantly extending the analytical depth compared to existing literature.

These methodological innovations—integrating subjective self-reported data with quantile regression—constitute the core novelty and contribution of this paper, providing enhanced insights into household-level herding behaviors in China.

This paper is organized as follows: In Section 2, we discuss the past literature regarding the herding effect based on different approaches. In Section 3, we introduce the data used and the structure of the quantile regression model. In Section 4, we interpret the empirical results of the quantile regression estimations along with the decomposed herding effect analysis. Section 5 includes the conclusion.

2. Literature Review

Herding can be more apparent when investors feel confident regarding the future direction of the market. Once a crisis appears, beta herding becomes much weaker, as a concern for fundamentals takes over (Hwang & Salmon, 2006). Macroeconomic information such as behavioral biases, demographics, and financial literacy, rather than firm-specific information, tends to have a more significant impact on investor behavior in markets which exhibit herding (E. C. Chang et al., 2000). The changes in stock investment returns can be caused by various factors. To comprehensively understand the issue, factors other than the herding effect have been inclusively discussed. Therefore, the literature regarding different factors influencing stock investment is reviewed in this section.

2.1. The Herding Effect on Stock Returns

In financial markets, herding behavior refers to the trading action taken by investors based on information received instead of their underlying fundamentals. Such behavior violates the principle of rational asset pricing, which is based on equity fundamentals (Saastamoinen, 2008). The herding effect can be identified by capturing the decrease in stock return dispersion or the increase in market return dispersion at a less-than-proportional rate (E. C. Chang et al., 2000). Measuring the dispersion of stock returns has become a mainstream method for observing the herding effect. As shown in early studies, detailed records of individual trading activity are required to identify the existence of herding

behavior. However, such data are not always readily available (Lakonishok et al., 1992; Wermers, 1999).

Modifying the traditional CAPM model and focusing on the beta value to examine the existence of the herding effect is another mainstream approach. In some advanced studies, the research methodology is based on the simple cross-sectional variability of returns (Christie & Huang, 1995; E. C. Chang et al., 2000); however, a method designed in such a way only reflects fundamental changes in common pricing factors, where the issue of irrational pricing cannot be captured. A new method was proposed to better examine the herding effect based on the cross-sectional variability of betas (Hwang & Salmon, 2006). Compared with earlier studies, the beta herding method captures irrational pricing more effectively, while other approaches are incapable of clearly differentiating irrational and rational pricing from the fundamentals.

Most empirical studies on herding behavior and its impacts have been conducted in developed market contexts, yet evidence suggests a stronger herding phenomenon in emerging markets due to weaker financial reporting requirements, lower accounting standards, insufficient regulatory enforcement, and higher information acquisition costs (Bikhchandani & Sharma, 2000). Studies specifically examining emerging markets also underline heightened susceptibility to behavioral biases, such as herding, driven by information asymmetry and regulatory weaknesses (Economou et al., 2018; Guney et al., 2017).

In the context of China, herding behavior has been found to be particularly prevalent due to the dominance of retail investors, limited financial literacy, and speculative trading behaviors (Tan et al., 2008; Yao et al., 2014).

2.2. How Debt Influences Stock Returns

In a previous study, the authors employed the portfolio choice model and discovered that households with debt mortgages have a 10% lower probability of being engaged in stock investment. Some households simply give up on stock investment due to interest payments on debt (Becker & Shabani, 2010). In another study, the authors used path analysis and indicated that there is no significant bidirectional relationship that can be found between investing in real estate and holding risky financial assets, as investment in real estate may generate alternative financial assets and hence decrease the need to hold risky financial assets (Chen & Song, 2021). These findings are beneficial to understanding the relationship and structure of household debt and investment portfolios. In this study, total debt is used to observe the influence of household debt on investment performance.

2.3. The Influence of Information Credibility

While the efficient market hypothesis (EMH) simply states that information is shared by all investors in the market, whether the information is correctly interpreted by investors is an issue that is commonly discussed in the field of behavioral finance. In fact, whether the information disclosed is credible or not is a subjective topic, and how investors utilize the information and take action is worthy of discussion. The overconfidence of investors in interpreting information may affect their actions on trading, while investors who are risk-averse are less likely to change their existing pattern of trading behavior (Sajid, 2015). Investors' risk preferences matter in decision-making; additionally, deciding whether or not to use financial tools to support investment is also influential. Furthermore, decent corporate governance should be considered important information, in addition to stock pricing (Sajid, 2015). However, sufficient financial knowledge and literacy are essential.

2.4. The Influence of Financial (Risk) Literacy

Every investment decision affects the outcome. Whether it is rational or irrational, individuals make decisions based on the information they have obtained, and correct

decisions are more likely to result in better outcomes. Individuals with decent financial literacy are more capable of making proper or suitable investments (Hilgert et al., 2003). Insufficient financial literacy usually leads to limitations for investors when it comes to making proper investment decisions; hence, financial literacy is believed to be positively correlated with investment return. If adults have accrued sufficient financial knowledge during their schooling, they usually have more savings and asset accumulation (Bernheim et al., 2001). Furthermore, financially educated families are more capable of diversifying their asset portfolios, and hence improving their investment efficiency (Campbell, 2006). Education is generally beneficial for improving returns from investment, financial management, and increasing credit scores (Cole et al., 2012). It is suggested that one's financial literacy can be significantly improved if receiving financial education, which enhances confidence in decision-making and financial behavior, such as pension preparation, cash investment, and credit management (Hira & Loibl, 2005). It also has been discovered that there is a connection between financial literacy and performance on stock investment (van Rooij et al., 2011).

2.5. The Influence of Demographics

On average, women exhibit stronger herding behavior than men because women are generally more risk-averse compared with men (Lin, 2011). Individuals with formal education (in any field) also exhibit higher levels of overconfidence, as they believe they know more than they actually do (Bhandari & Deaves, 2006). Nevertheless, behavioral biases have a greater influence on herding behavior than cultural factors, and it has been confirmed that excessive optimism leads to a higher tendency to herd (C. Chang & Lin, 2015) and that inexperienced and younger investors are more likely to herd than older and more experienced investors (Lin, 2011). In another study, no gender-based herding effects were found, but moderate herding behavior was observed among those under 26 and those over 55 (Talpsepp & Tänav, 2021). Therefore, age was noted to have a significant negative correlation with herding behavior among stock market investors. Further, individuals with economics degrees tended to herd less than those with other degrees, suggesting that stronger financial knowledge can reduce herding behavior in the real estate market. Other researchers did not find evidence supporting the hypothesis that higher education or cognitive ability leads to lower levels of herding. Finally, the results of another study suggested that age and income are negatively correlated with herding behavior among investors; as age and income increase, herding behavior among individual investors decreases. Gender was not found to be a significant factor (Sabir et al., 2020). As discussed above, there are conflicts among studies: some research indicates that age and income have different correlations with herding behavior, while others suggest that the impacts of age and income on herding behavior differ. Therefore, there are studies that provide systematic support for the mediating role of herding in amplifying market sentiment, which leads to a crash risk (Bashir et al., 2024).

2.6. Hypothesis Development

Based on the literature discussed above, we have developed specific hypotheses regarding the determinants and consequences of household herding behavior in stock investment.

Hypothesis 1. *Household herding behavior is negatively associated with stock investment returns. Previous studies indicate that herding often leads to irrational investment decisions, mispricing, and subsequent lower returns (Christie & Huang, 1995; Yao et al., 2014). We hypothesize that households exhibiting stronger self-reported herding tendencies experience worse investment outcomes.*

Hypothesis 2. *The negative effect of herding on stock investment returns varies across the quantiles of returns. Given that investors differ in terms of financial literacy, risk tolerance, and information accessibility, we posit that herding behavior's negative impact is heterogeneous across different quantiles of the investment return distribution, highlighting the importance of using quantile regression analysis.*

Hypothesis 3. *Household demographic and financial characteristics moderate the relationship between herding behavior and stock investment outcomes. Drawing from prior research (Lin, 2011; Talpsepp & Tänav, 2021), we expect that demographic factors (age, education, CCP membership, gender) and financial factors (income, assets, debt, liquidity) interact with herding behavior, influencing its effect on stock returns and ROI differently.*

Hypothesis 4. *Information credibility perceived by investors moderates the negative relationship between herding behavior and investment performance. Consistent with behavioral finance literature, credible financial information can mitigate irrational investment decisions, suggesting a moderating role of perceived information credibility (Sajid, 2015).*

These hypotheses will be systematically tested using the quantile regression and quantile regression decomposition methodologies outlined in Section 3.

3. Methodology

3.1. Data Resource

The data in this study were gathered from the fourth round of the China Household Finance Survey (CHFS) in 2017, which was originally completed by the Research Centre for China Household Finance at Southwestern University of Finance and Economics. The data samples cover 40,011 households across different administrative regions, including 29 province-level regions and municipalities, 355 prefecture-level regions and sub-provincial cities, and 1428 village-level regions. The contents of the survey include household finance information regarding assets, liabilities, incomes, consumption, insurance, and personal characteristics, such as employment status and payment habits. The CHFS contributes to academic research and government policy design by offering detailed information on household finance and economics at the micro-level. The accuracy of statistical analysis results in representing the true situation of the population depends on several factors: the randomness of the sample, the correctness of the model, the absence of calculation errors, and the appropriateness of the method used for model analysis. Any errors in these aspects can lead to misleading results. Typically, due to budget and time constraints, samples represent only a small fraction of the total population. The China Household Finance Survey (CHFS) employed a stratified, three-stage, and probability-proportional-to-size (PPS) sampling design. The primary sampling units (PSUs) comprised 2585 counties and cities. In the second stage, neighborhoods/villages were sampled directly from these counties/cities. In the final stage, households were sampled within these neighborhoods/villages. The response rate was approximately 88.4%. The distribution of per capita GDP was designed to be consistent with the overall population rather than geographical distribution, which may have led to discrepancies in geographical representation.

The data were filtered and processed two stages prior to numerical analyses to maintain accuracy in accordance with the study goal. In the first stage, only the heads of the household were selected as valid respondents, and their corresponding responses were used as the dependent variables in this study, including age, gender, education level, Chinese Communist Party (CCP) membership, and most importantly, herding behavior. In the second stage, the data were further filtered, retaining only the responses that provided

household information regarding capital and stock investment results selected as valid. Among all the selected data, if any value was either missing or deemed unreasonable due to respondents' negativity and dishonesty, the corresponding samples were excluded. A total of 1400 household data points were selected as the final samples, and the data were numerically recoded and standardized to maintain quantitative consistency.

3.2. Statistical Model and Variables

Traditional linear models explain the dependent variable y as a continuous variable, producing a single coefficient value for each independent variable x , which fails to capture different effects that x may have at various quantiles of y . If the probability of y being less than or equal to a certain value is τ , we refer to q_τ as the quantile of y . When dependent variable Y is expressed as a linear function of a series of explanatory variables X , satisfying the condition that the probability of y being less than or equal to this function is q_τ , this is known as quantile regression.

The shortcomings of the traditional linear models and the least squares method in data analysis, such as the tail effect, extreme values, and the occurrence of sudden events, can be addressed with quantile regression. For the model characteristics where multiple distributions are assumed individually at each quantile, the aforementioned shortcomings can be mitigated, leading to more accurate estimations. The quantile regression model was proposed to differentiate research objects into different quantiles, each representing the distributional uniqueness of samples at different levels (Koenker & Bassett, 1978). The coefficient estimators are often inconsistent, indicating that influencing factors have different effects on the research objects at different levels. In particular, when the distribution of the research samples is heterogeneous, such as in cases of asymmetry, thick tails, truncation, and deviation, this method provides robustness to outliers in the data; furthermore, quantile regression makes no distributional assumptions on random perturbation, and the entire regression model is robust (Koenker & Hallock, 2001; Y. Wu & Liu, 2009). Consequently, the quantile regression method is commonly adopted for stock market analyses (Barnes & Hughes, 2002; Chiang & Li, 2012; Ma et al., 2018). Quantile regression allows us to accurately observe the effects of independent variables at different quantiles of the dependent variable, thus providing multiple coefficient values and revealing various quantile effects. This approach overcomes the limitation of traditional regression, which yields only a single coefficient.

With the quantile regression model, an application where multiple least squares methods assume the estimation of coefficients follows an asymptotic normal distribution. The distribution function of the random variable is written as $F(y) = P(Y < y)$. Variable q determines a random sample set of $Y = \{y_1, y_2, \dots, y_n\}$ with respect to the q th percentile. By summing up the positive and negative errors, the optimal values (the minimized ones) lead to proper estimated coefficients at the corresponding quantile. The minimization process is described as follows:

$$\min \left\{ \sum_{i: y \geq x_i' \beta(q)} q |y_i - x_i' \beta(q)| + \sum_{i: y < x_i' \beta(q)} (1 - q) |y_i - x_i' \beta(q)| \right\} \quad (1)$$

where x'_i is the vector of dependent variables and $\beta(q)$ is the coefficient vector, which varies for different q . The linear conditional quantile function is written as follows:

$$Q_\tau(y_i | x_i) = x_i' \beta(q) \quad (2)$$

For any τ between 0 and 1, the coefficient $\beta(q)$ indicates the corresponding q th regression quantile. To directly estimate the asymptotic covariance matrix across every $\beta(q)$, the bootstrapping method is adopted following an early study by (Chuang & Kuan,

2005). In this study, the explained variable (Y) denotes the returns earned from household stock investment.

3.3. Description of Variables

3.3.1. Basic Information on Head of Household

The variables adopted in this study were the characteristic factors describing the socioeconomic status of the head of the household, including gender, age, education level, CCP status, sense of social trust, and sense of happiness, which were selected as the observation variables of the family's socioeconomic status. The dummy variable was employed to indicate the differences in gender and CCP status. In either of the responses, 1 denotes male CCP members; in contrast, 0 denotes female non-CCP members. The education level was divided into nine categories according to the ranking and measurement of the questionnaire, processed as ordinal variables. The variables for assets, income, debt, deposits, and cash were measured in units of RMB ten thousand. The variables measuring herding effect, social trust, and happiness were evaluated on a scale from 1 to 5, with higher numbers indicating a stronger degree of each respective variable.

3.3.2. Head of Household's Financial Status

The variables representing households' financial status included annual assets, income, total debt, deposits, and cash spending. All empirical values were divided by 10,000 RMB. If any value was missing, the corresponding sample was deleted.

3.3.3. Other Variables

The credibility of the information disclosed by listed companies to the head of the household was measured on a scale of 1–5, with higher scores indicating higher credibility. The herding effect variable was measured based on subjective self-reported data from the head of the household, based on question D3111d in the CHFS: "Are you susceptible to comments (behaviors) by other investors around you?" A score of 5 suggests that these heads of the household admitted to being extremely susceptible when investing, and a score of 1 suggests they displayed conformity to their own judgment. The variable of risk literacy was obtained by adding up the answers to two questions, H3103 and H3115, with 1 point for each correct answer, and a maximum total score of 2 points. Adopting the above-mentioned variables provides valuable individual-level insights, although it may introduce biases arising from respondents' subjective self-assessments; however, the large number of variables included in this study, ranging from demographic characteristics to financial status, behavioral traits, and attitudinal responses, helps reduce omitted variable bias and strengthens the robustness of the analysis. By capturing a wide spectrum of household-level factors, the model can better isolate the effect of herding behavior and mitigate confounding influences to some extent.

3.3.4. The Dependent Variable—The Stock Investment Return of the Head of Household

In this study, we included two dependent variables to describe the stock investment return of the head of the household focusing on unrealized and realized gains and losses. Unrealized gains or losses are defined as the current market value of stocks held by the head of the household minus their holding cost (values were provided by the household head). Realized gains or losses were derived from either stock trading profits or dividends received in the past year, and the values were also provided directly by the head of the household. The first dependent variable was the stock investment return, which we calculated by summing the unrealized and realized gains and losses (divided by RMB ten thousand). The second dependent variable was the stock return on investment (ROI) of the head of the

household in the current year, which was calculated by dividing the total gains and losses by the holding cost.

3.4. Operational Definitions of Variables

For clarity and replication purposes, we provide explicit operational definitions for each variable utilized in our analysis.

- **Dependent Variables**

Stock Investment Return: The sum of unrealized and realized stock gains or losses reported by households, measured in RMB ten thousand. **Stock return on investment (ROI):** The ratio of total gains and losses (realized plus unrealized) to the stock holding cost reported by households, expressed as a decimal.

- **Independent Variables—Household Head Characteristics**

Gender: A binary variable coded as 1 for male and 0 for female. **Age:** Continuous variable measured in years. **Education level:** Ordinal variable, measured from 1 (no formal education) to 9 (doctoral degree), based on respondents' self-reported highest education attainment. **CCP Membership:** A binary variable, coded as 1 if the household head is a member of the Chinese Communist Party (CCP), and 0 otherwise.

- **Independent Variables—Household Financial Characteristics**

Asset: Continuous variable representing total household assets, measured in RMB ten thousand. **Income:** Annual total household income, measured in RMB ten thousand. **Debt:** Total household outstanding debt, measured in RMB ten thousand. **Deposit:** Total household deposits held in banks, measured in RMB ten thousand, representing household liquidity. **Cash holding:** Amount of cash held by households, measured in RMB ten thousand.

- **Attitudinal and Behavioral Variables**

Information credibility: Respondent's perception of the credibility of financial information disclosed by listed companies, measured on a 5-point Likert scale (1 = very low credibility; 5 = very high credibility). **Herding effect:** Self-reported tendency of household heads to be influenced by other investors' comments or behaviors, measured on a 5-point Likert scale (1 = strongly independent; 5 = strongly influenced). **Sense of social trust:** Respondent's reported level of trust in society, measured on a 5-point Likert scale (1 = very low trust, 5 = very high trust). **Sense of happiness:** Respondent's subjective perception of happiness, measured on a 5-point Likert scale (1 = very unhappy, 5 = very happy). **Risk literacy:** Respondent's level of financial risk literacy, measured by summing correct answers (0–2) to two specific financial literacy questions (H3103 and H3115 in CHFS).

These definitions facilitate clear understanding and consistent interpretation of the empirical analyses presented in this paper.

3.5. Addressing Potential Endogeneity

In this study, we recognize potential endogeneity concerns that commonly arise in observational analyses, such as omitted variables, simultaneity, and measurement errors, particularly given the self-reported nature of our herding variable. To proactively address these concerns, we have implemented several methodological precautions. Firstly, our comprehensive inclusion of demographic, socioeconomic, financial, and attitudinal control variables significantly reduces the likelihood of omitted variable bias. Secondly, by employing quantile regression methods, we effectively capture the heterogeneous impacts across different investor segments, thereby mitigating issues related to unobserved heterogeneity and providing robust insights into the differential effects of herding behavior. While instrumental variable approaches are beyond the scope of our current dataset, our

rigorous analytical approach enhances confidence in the reliability of the presented results. Future studies incorporating additional methodological refinements, such as instrumental variables or experimental approaches, would further enrich our understanding of the causal dynamics of herding behavior in investment decisions.

3.6. Descriptive Statistics

The descriptive statistical values of each variable are shown in Table 1, while Table 2 presents the return distributions of the two dependent variables. It is notable that the stock investment return shifts from negative to positive at the 60th quantile in both measures.

Table 1. Descriptive statistics.

Variables	Min.	Max.	Mean.	S.D.	Obs.
Gender	0	1	0.67	0.469	1400
Education level	1	9	5.33	1.663	1400
Age	18	90	50.81	13.393	1400
CCP membership	0	1	0.79	0.409	1375
Asset	1.329	3000	349.628	412.059	1400
Income	−1000	2534.500	193.042	242.116	1400
Debt	0	5000	145.907	456.202	1400
Deposit	0	3000	144.803	269.713	1400
Cash holding	0	2000	16.009	67.370	1400
Information credibility	1	5	2.25	0.958	1355
Herding effect	1	5	2.24	1.232	1384
Sense of trust	1	5	2.44	0.873	1394
Sense of happiness	1	5	3.78	0.734	1396
Risk literacy	0	2	1.62	0.575	1400
Return amount	−175	337.628	−1.062	22.867	1400
ROI	−7.021	1246.200	1.613	33.886	1400

Table 2. Distributions of the two dependent variables.

	Return	ROI
10	−120,000	−0.7917
20	−60,000	−0.5825
30	−30,092.4	−0.4285
40	−15,000	−0.2954
50	−4000	−0.1299
60	0	0
70	5948.1	0.1219
80	30,000	0.4637
90	100,000	1.4168

4. Empirical Results

The independent variables representing the basic information of the head of the household in this study included gender, education level, age, and CCP membership. The independent variables for financial status were assets, income, debt, deposits, and cash holdings. Additionally, the information credibility of listed companies, the herding effect, sense of social trust, sense of happiness, risk literacy, and 14 other variables were considered. It is necessary to verify whether a collinearity problem exists among these independent variables. Therefore, the OLS method was used to examine the variance inflation factor (VIF) of the independent variables. The values of the VIF ranged between 1.032 and 1.615 (the tolerance values ranged from 0.619 to 0.969), indicating that no severe collinearity was

detected among the respective variables. Subsequently, the quantile regression of the two dependent variables could proceed.

To enhance the robustness of the quantile regression results, we employed a non-parametric bootstrap procedure with 1000 replications to compute robust standard errors and confidence intervals. The results showed that the estimated standard errors were consistently small, indicating a high level of estimator stability across resampled datasets. A summary of these estimates is provided below in Table 3.

Table 3. Bootstrap-based robust standard errors from quantile regression.

	Coef.	Std. Err.	t	$p > t $	[95% Conf. Interval]	
Gender	0.011515	0.248696	0.05	0.963	−0.47638	0.499406
Education	−0.02297	0.072142	−0.32	0.75	−0.1645	0.11856
Age	0.001152	0.00858	0.13	0.893	−0.01568	0.017984
CCP	0.535608	0.311956	1.72	0.086	−0.07639	1.147603
Asset	−0.00013	0.000627	−0.21	0.83	−0.00137	0.001096
Income	0.002545	0.001691	1.5	0.133	−0.00077	0.005862
Debt	−0.00056	0.000586	−0.96	0.337	−0.00171	0.000588
Deposit	-2.7×10^{-5}	0.001446	−0.02	0.985	−0.00286	0.00281
Cash	3.61×10^{-6}	9.97×10^{-6}	0.36	0.718	-1.6×10^{-5}	2.32×10^{-5}
Information	0.640249	0.127808	5.01	0	0.389516	0.890982
Herding	−0.44753	0.116051	−3.86	0	−0.6752	−0.21986
Trust	−0.21364	0.142298	−1.5	0.134	−0.4928	0.065522
Happiness	−0.33557	0.158255	−2.12	0.034	−0.64603	−0.02511
Risk literacy	−0.26536	0.197274	−1.35	0.179	−0.65237	0.121654

In addition, Wald-type tests were conducted to assess whether the estimated coefficients differed significantly across quantiles. The results, as shown in Table 4 below, confirmed that several key predictors (e.g., asset level, age, income, and information credibility) exhibited statistically significant differences across quantiles, supporting the existence of distributional heterogeneity and justifying the use of quantile regression.

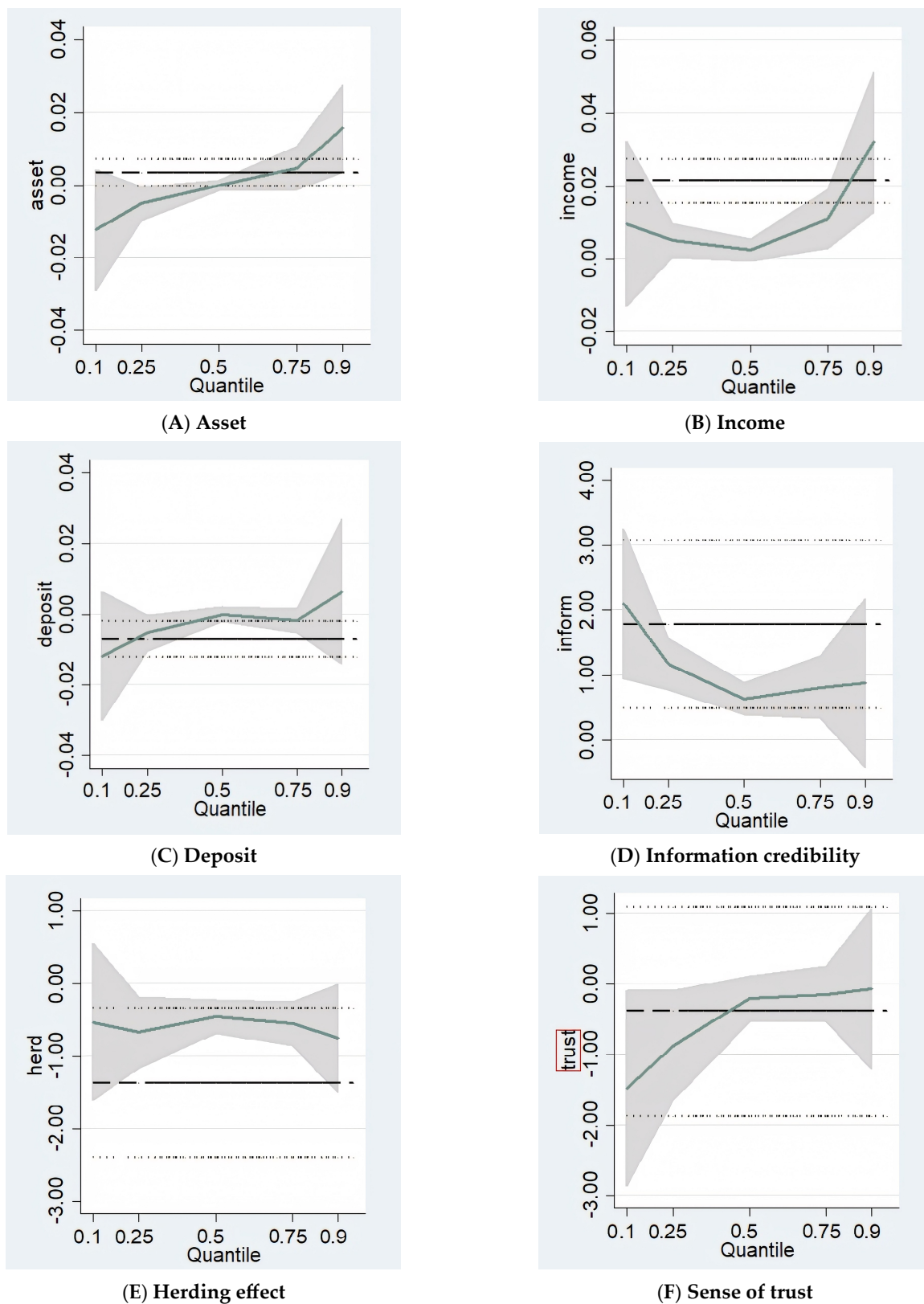
Table 4. Wald tests for coefficient equality across quantiles ($q_{25} = q_{50} = q_{75}$).

Variables	F-Value
Age	2.19 *
Asset	6.97 ***
Income	2.86 *
Information credibility	2.54 *
Risk literacy	1.79 *

*** $p < 0.01$; * $p < 0.1$.

4.1. Total Stock Investment Returns: Summing Unrealized and Realized Gains and Losses

By taking the total amount of realized and unrealized stock investment gains and losses as the dependent variables to observe their relationships with the respective variables, the actual profitability can be accurately presented because the hidden social desirability (where the realized gains are zero) is not considered. Table 5 presents the quantile estimation results of q_{10} , q_{25} , q_{50} , q_{75} , and q_{90} for all independent variables. The lower the quantile value, the lower the stock investment return of the head of the household; conversely, the higher the quantile value, the higher the stock investment return of the head of the household. Figure 1 illustrates the quantile changes in 10 variables, namely, assets, income, deposits, age, education level, CCP membership, disclosed information credibility, herding effect, sense of social trust, and risk literacy. The coefficient values vary significantly at different quantiles.



Note: The solid line represents the estimated coefficient at each quantile, and the shaded area indicates the 95% confidence interval.

Figure 1. Coefficient trend across all quantiles (stock investment return).

Table 5. Quantile regression results on stock investment returns.

	q10	q25	q50	q75	q90
Gender	0.976	−0.019	0.012	−0.025	−1.06
Education level	−0.074	−0.105	−0.023	−0.312 *	−0.277
Age	−0.033	−0.029	0.001	0.015	0.1162 ***
CCP membership	6.0052 *	1.209	0.5362 *	−0.118	−0.682
Asset	−0.0122 **	−0.0052 **	−0.0001	0.0052 **	0.0162 **
Income	0.01	0.0052 *	0.003	0.0112 **	0.0322 ***
Debt	−0.004	−0.001	−0.001	−0.002	−0.003
Deposit	−0.012	−0.0052 *	−0.00003	−0.002	0.006
Cash holding	−0.024	−0.014	0.004	0.021	−0.006
Information credibility	2.1122 ***	1.1742 ***	0.642 ***	0.8182 ***	0.8842 *
Herding effect	−0.525	−0.6662 **	−0.4482 ***	−0.5462 ***	−0.7432 *
Sense of trust	−1.4832 *	−0.8742 **	−0.214	−0.147	−0.069
Sense of happiness	−0.97	−0.366	−0.3362 **	−0.194	−0.352
Risk literacy	−2.1052 ***	−0.924	−0.265	−0.46	−0.294
Constant	−5.08	1.405	0.417	1.518	−1.618

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

According to Table 5, regarding the demographic variables, the education level negatively affects stock investment returns only at the 75th quantile. Regarding age, an increasingly positive effect on stock investment returns can be observed starting from the 90th quantile. CCP membership has significant and positive coefficients at the 10th and 50th quantiles, indicating that the heads of the household with CCP membership tend to achieve better stock investment results. No significant differences were found for the gender variable.

Regarding the financial variables, the variable of assets is significantly correlated with stock investment returns across all quantiles except for q50, with coefficient values exhibiting volatility across these quantiles. The influence of assets on stock investment efficiency shifts from negative to positive as the quantile increases, transitioning from a concave to a convex pattern, as depicted in Figure 1A. This indicates a turning point around q60, where the variable coefficient becomes positive. Compared with the results shown in Table 2, positive stock investment returns are observed only after q60. For heads of the household in the q10–q50 range, a relationship between stock investment loss and asset reduction is evident. Nonetheless, the increasing trend of the coefficient indicates that assets still exert a positive influence on stock investment returns. These complex but comprehensive findings concerning a single variable underscore the advantages of the quantile regression model, which provides a range of coefficient values rather than a single average value. The income coefficients are positive across all quantiles, with income having a significant positive effect on stock investment returns above q75. Similar to the impact of assets, the effect of income on stock investment returns shows a positive increase starting from q25, where a reversal effect is noticeable. The deposit amount, which measures the effect of capital liquidity, shows negative coefficients before q75, with a significant negative correlation observed only at q25. At q25, the significant negative correlation indicates a reduction relationship between capital liquidity and stock investment returns, meaning that investment failures indeed shrink investors' capital liquidity. However, the increasing trend across quantiles suggests that the deposit variable also exerts a positive influence. Other financial variables showed no significant effect.

In terms of various attitudinal variables, the credibility of the information disclosed by listed companies to the head of the household shows significant and positive estimation results across all quantiles. However, significant inconsistencies are evident. Before the 50th quantile, the disclosed information results in stock investment losses. For heads of the

household above the 75th quantile, increased trust in the disclosed information correlates with greater investment. It is possible that, as minority information is acquired, excessive investment returns can be easily realized. The senses of social trust and happiness held by the head of the household are negatively correlated with stock investment returns, although this trend is increasing. Before the 50th quantile, a higher sense of social trust and happiness benefits increased stock investment returns. The negative relationship between social trust and investment returns observed at lower quantiles may reflect risk-averse behavioral tendencies. However, given the absence of consistent effects across quantiles and the limited theoretical consensus on this mechanism, we recommend interpreting these results with caution. Similarly, the negative coefficients for self-reported happiness at certain quantiles may be confounded by unobserved psychological or emotional factors. Risk literacy shows a statistically significant negative effect only at the 10th quantile; however, this effect lies close to the significance threshold. At other quantiles, the estimates remain negative but are not statistically significant. These results suggest that among households experiencing the most substantial investment losses, even higher self-assessed risk awareness does not effectively prevent poor investment outcomes, implying that professional knowledge is essential to successful stock investment. Nonetheless, given the marginal significance, these findings should be interpreted with caution and verified with future studies.

The herding effect, which is subjective self-reported data, particularly regarding the head of the household, shows a significant negative correlation with stock investment returns, except at q10. However, inconsistencies across different coefficients, including rapid decreases at higher quantiles, provide varying interpretations. Up to q50, the heads of the household experience stock investment losses, indicating that reducing herding behavior can decrease losses, though it does not necessarily guarantee increased profitability. For heads of the household above q50, there is a strong reverse effect, suggesting that the absence of herding behavior is crucial to enhancing stock investment profitability. The herding effect, which is based on subjective self-reported data, particularly regarding the head of the household, shows a significant negative correlation with stock investment returns, except at q10. However, this subjective measure could introduce bias due to individual perception differences.

In summary, for heads of the household, the herding effect is detrimental, and independent judgment on stock investment opportunities tends to be more profitable. However, without sufficient professional knowledge, such independent decisions may not yield profits. Based on these findings, it can be inferred that the herding effect influences different investment groups to varying degrees, depending on whether they are gaining or losing at different quantiles. To fully understand this inference, it is essential to consider the simultaneous observations of other variables rather than focusing solely on one independent influencing factor. This underscores the motivation for adopting the quantile regression model for analysis in this study, a topic that is critical in behavioral economics. The herding effect does not impact everyone equally, and its degree may vary among different groups.

4.2. Realized and Unrealized ROI as Dependent Variables

In the empirical results, with ROI as the dependent variable, when suffering a loss (q10~q50, as shown in Table 6), the male head of the household seems to experience fewer losses than the female head of the household at q10. For the age variable, the heads of household at q50 and q90 achieve better ROI, suggesting that increasing ROI is more easily achieved due to better and longer investment experience. This pattern is more prominent at higher quantiles; the ROI of the elderly increases and rises sharply after q80. For CCP members, the pattern is similar to that of age, with a significant improvement in ROI at q50 and a sharp rise after q80.

Table 6. Quantile regression results on ROI.

	q10	q25	q50	q75	q90
Gender	0.096 **	0.049	0.035	−0.062	0.083
Education level	0.017	0.023	−0.007	−0.033	−0.081
Age	0.003	0.002	0.002 *	0.004	0.03 ***
CCP membership	0.011	0.037	0.081 *	0.096	0.0004
Asset	3.87×10^{-5}	0.00004	4.58×10^{-5}	0.0003	0.001
Income	0.0002 **	0.0002 **	0.0003 *	0.0009 ***	0.001
Debt	-7.24×10^{-6}	-6.6×10^{-5}	-4.1×10^{-5}	−0.00013	6.88×10^{-5}
Deposit	−0.0001	3.71×10^{-5}	5.99×10^{-5}	−0.00011	−0.0006
Cash holding	0.0003	0.0003	-4.3×10^{-5}	−0.0006	−0.0011
Information credibility	0.022	0.08 ***	0.12 ***	0.124 ***	0.256 **
Herding effect	−0.043 **	−0.052 ***	−0.063 ***	−0.106 ***	−0.082
Sense of trust	−0.009	−0.023	−0.032	−0.048	−0.093
Sense of happiness	−0.025	−0.038	−0.035	−0.005	0.103
Risk literacy	−0.003	−0.057 *	−0.045	−0.034	−0.102
Constant	−0.947	−0.605	−0.226	0.161	−0.67

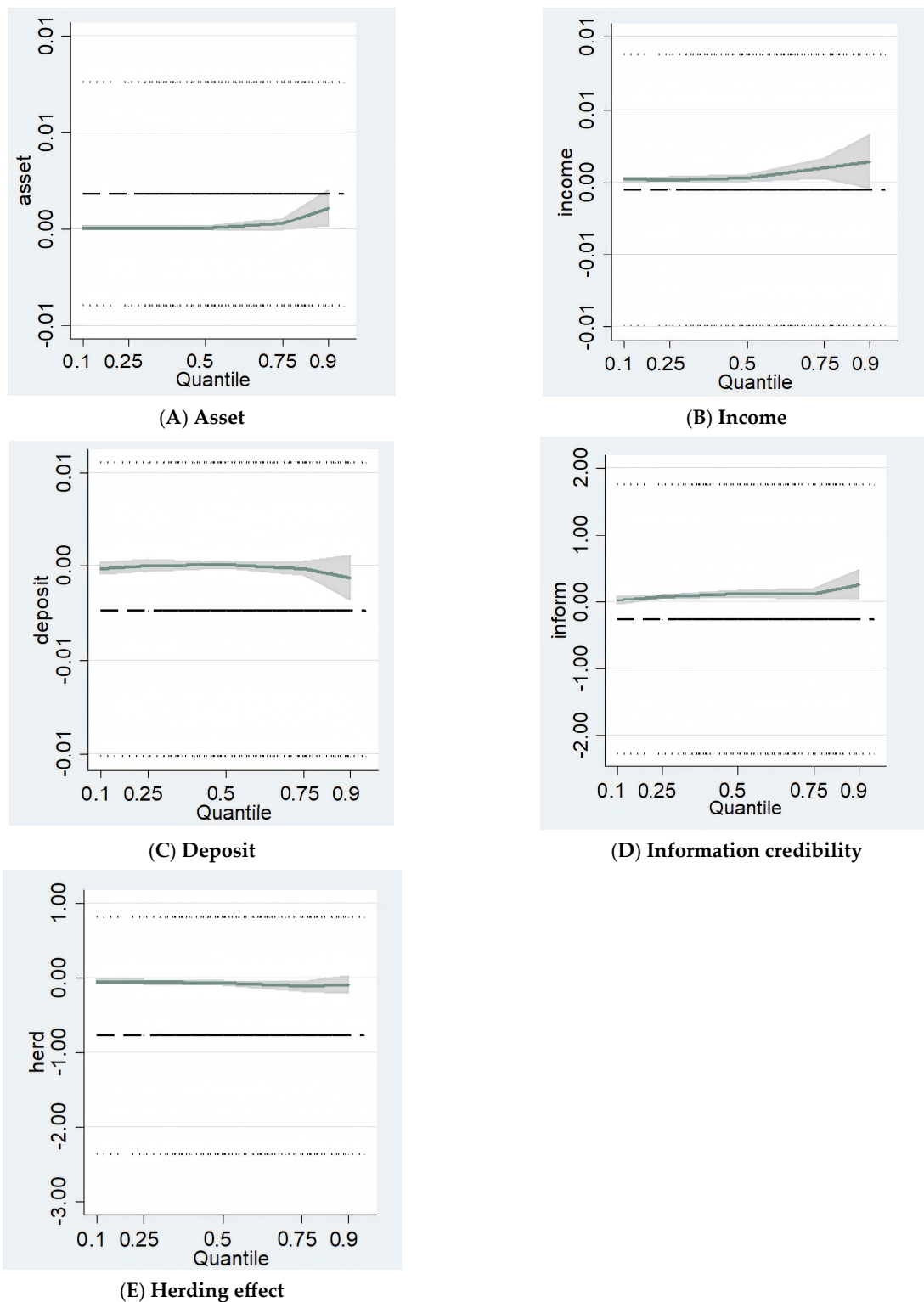
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Continuing to observe financial variables, it is noted that unlike the previous estimation result, according to which assets significantly influence the investment return, no significant correlation can be observed between assets and ROI. On the other hand, income is significantly and positively correlated with ROI. Regardless of whether investors are gaining (q70 and above) or losing (q10 to q50), income positively affects ROI, and this effect continues to increase until q80, as shown in Table 6. The sudden drop after q80 suggests that while investment increases as income increases, once income reaches a certain level, investment becomes less attractive to household investors. As for debt, as in the previous section, no significant effect on ROI can be observed.

Shifting our focus to attitudinal variables, the belief that financial information can reduce investment losses positively influences ROI. The credibility of the information disclosed by listed companies also shows a positive effect. If the financial information of a company is correctly interpreted, investment efficiency increases. The aforementioned stock investment return presents a U-shape and rises after q80 (as shown in Figure 2D), indicating a strong connection between trust in financial information and the increase in profitability. Additionally, as shown in this section, the correlation between information credibility and ROI increases positively with higher quantiles. The estimation results regarding social trust and happiness are consistent with those in the previous section, but no significant correlation is observed. Interestingly, risk literacy exhibits a statistically significant negative relationship with ROI at the 25th quantile. This suggests that, under certain conditions, increased risk awareness may not directly translate into improved investment outcomes. However, as the significance level is marginal and the direction of the effect contradicts intuitive expectations, this result should be interpreted with caution. Further validation using alternative datasets or modeling approaches is recommended.

Focusing on the herding effect, it is negatively correlated with ROI, except at q90, which is consistent with previous estimation results regarding stock investment returns. For the head of the household with investment losses, reducing herding behavior is beneficial for reducing negative ROI; however, profits cannot be guaranteed with this action. For the head of the household above q50 (not losing), it can be suggested that herding behavior should be avoided to maintain a positive ROI. The empirical results for the two dependent variables differ significantly in terms of quantile changes. While the stock investment return shows a generally decreasing trend, ROI shows a decreasing trend in the first half

and then an increasing trend in the second half, indicating that the influence of the herding effect weakens as ROI increases.



Note: The solid line represents the estimated coefficient at each quantile, and the shaded area indicates the 95% confidence interval.

Figure 2. Coefficient trend across all quantiles (ROI).

Whether considering the return or ROI models, the education level, age, and CCP membership variables produce similar results, with positive effects observed for both asset and income variables. However, in the ROI model, the impact of assets is no longer

significant, indicating that investment returns are not related to assets, suggesting that other determining factors are more important. Information credibility consistently shows positive results, while the herding effect consistently demonstrates negative results.

4.3. Quantile Regression Decomposition

In most of the econometric literature, binary effect estimations focus primarily on the average treatment effect under exogenous influences. In this study, the concept of adopting quantile regression decomposition is proposed to replace the traditional mean reversion. In this model, by decomposing the variable gap of binary groups (such as identity, urban and rural areas, etc.) into the sum of the effects of various factors, the researcher can observe the effect of the gap (Oaxaca, 1973). For example, the definition of the wage difference between men and women is

$$D = \left[\left(W_m / W_f \right) - \left(W_m / W_f \right)^0 \right] / \left(W_m / W_f \right)^0 \quad (3)$$

where D is the discrimination coefficient measuring the degree of discrimination, $\left(W_m / W_f \right)$ measures the gap in actual wages between males and females, and $\left(W_m / W_f \right)^0$ measures the gap in actual wages between males and females under the assumption that discrimination does not exist. If discrimination does not exist, the wage structure for both males and females should be the same. As the estimated coefficient matrices of men and women have been calculated, the wage difference is observable by the subtraction of both and can be further disassembled into explainable (discrimination) and unexplained differences.

From both quantile regression estimation procedures in the earlier sections, it can be observed that the herding effect is generally significant regardless of the dependent variable, and the coefficients are all negative. Therefore, conducting a quantile regression decomposition on this specific independent variable for a better understanding of the underlying mechanisms is considered worthwhile and reasonable. By adopting the proposed quantile regression decomposition model (Melly, 2005, 2006), the return differences were decomposed, with the herding effect as the grouping variable. In this model, a binary group is determined by using a dummy variable, where the value is 1 if the head of the household is easily influenced when making evaluations and selections; otherwise, it is 0. In this study, we rewrite the formula describing the differences in household stock investment returns as follows:

$$SR_q^1 - SR_q^0 = \left(X^1 \beta^1 - X^1 \beta^0 \right) + \left(X^1 \beta^0 - X^0 \beta^0 \right) \quad (4)$$

SR_q^1 is the average stock investment return of the head of the household in the τ quantile with herding behavior; SR_q^0 is the one without herding behavior; and $SR_q^1 - SR_q^0 = (1,0)$ determines the gap of average stock investment return of the two groups. Within the estimated gap, as β^0 and β^1 are the OLS estimation coefficients, $X^1 \beta^0$ is a counterfactual distribution, where the statistical coefficients of the group without herding behavior are taken to estimate the observation values of the group with herding behavior, namely, $F(SR_q^1 | X^1 \beta^0)$. The first term on the right of Equation (4) is the difference in the return distribution caused by the herding effect (coefficient effect) and is generally regarded as a discriminatory difference; the second term is called the characteristic effect, which represents the difference in the distribution of returns due to differences in individual characteristics. In this study, the logic of the discriminatory effect is adapted to analyze the herding effect due to the variable's binary attribute.

4.3.1. Decomposing the Impact of Herding Behavior on Stock Investment Returns

Table 7 presents the decomposition results of the differences in stock investment returns, alongside the ratios of the herding effect and the characteristic effect across quantiles. The second column on the left of Table 7 represents the difference in average stock investment returns between the two groups, with and without the herding effect ($SR_q^1 - SR_q^0$), and is then decomposed into characteristic differences and herding differences. Since the differences in stock investment returns are all negative values from q10 to q90, this indicates that the heads of the household exhibiting herding behavior have poor stock investment performance. An inverted U-shaped trend (as shown in Figure 3 with a solid line) is formed, with values ranging from -5.99 at q10 to -1.967 at q60 and then gradually decreasing to -6.016 at q90. From q10 to q90 (from the low-investment-return group to the high-investment-return group, respectively), the herding difference exists at every quantile (as shown in Figure 3 with a dotted line). Furthermore, all negative values suggest that the herding effect is certainly unfavorable. The inverted U-shaped trend (as shown in Figure 4 with an interrupted line) demonstrates that the unfavourability is more pronounced at the higher quantiles (q80 and q90). The proportions of the herding difference to the stock investment return difference ranges from 55% to 95%. Therefore, significant influence by herding behavior on returns can be easily observed.

Table 7. Decomposition results of differences in stock investment return.

Quantile	Differences in Return	Characteristic Differences	Herding Differences	Ratios of Herding Effect
q10	−5.99	−2.635	−3.365	0.561
q20	−3.774	−1.317	−2.457	0.651
q30	−3.792	−0.816	−2.976	0.785
q40	−3.537	−0.492	−3.045	0.861
q50	−3.039	−0.315	−2.724	0.896
q60	−2.434	−0.172	−2.262	0.929
q70	−1.976	−0.263	−1.714	0.867
q80	−3.029	−0.436	−2.594	0.856
q90	−6.061	−0.289	−5.772	0.952

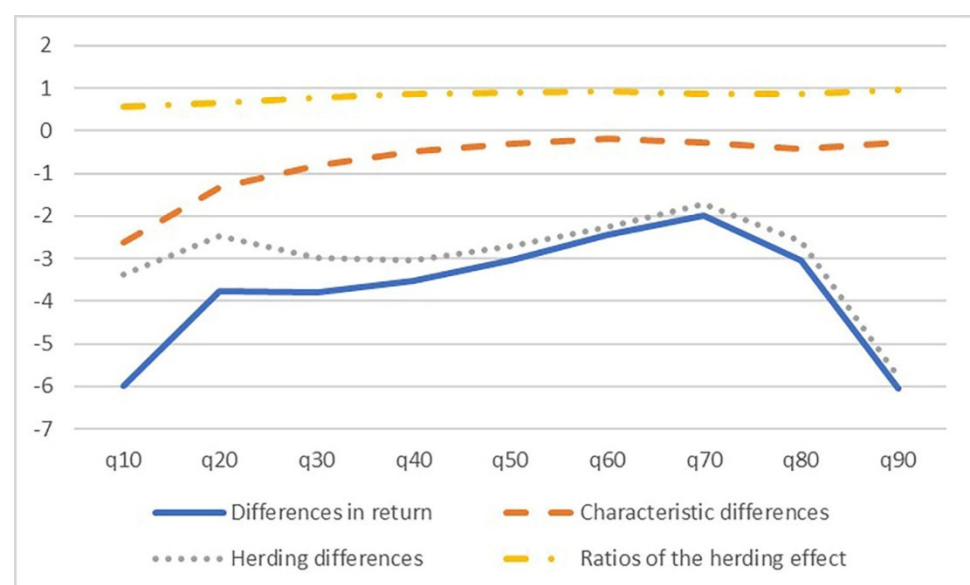


Figure 3. Decomposed differences regarding stock investment return.

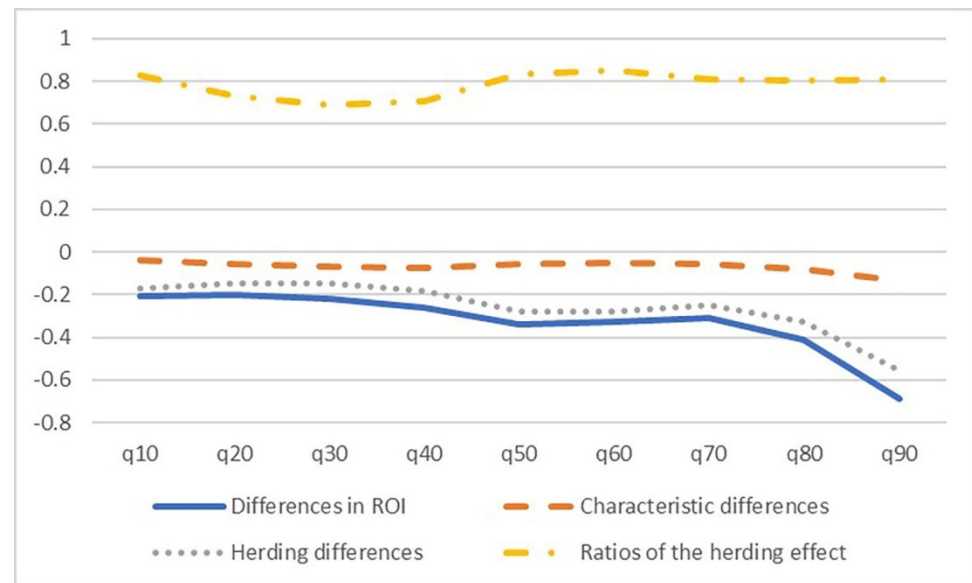


Figure 4. Decomposed differences regarding ROI.

4.3.2. Decomposing the Herding Effect on ROI

Table 8 presents the decomposition results of the difference in ROI, along with the ratios of the herding effect and the characteristic effect across quantiles. Similarly, the second column on the left represents the difference in ROI between the two groups, with and without the herding effect ($SR_q^1 - SR_q^0$), and is then decomposed into characteristic differences and herding differences. The ROI difference negatively increases in value from q10 to q90 (as shown in Figure 4 with a solid line), indicating that the herding effect is unfavorable for investments. For the purpose of increasing profitability, professional knowledge is necessary. As shown in Figure 4, the trend of the herding effect differences fluctuates along with the differences in ROI, indicating that the herding difference is the main factor influencing the ROI difference. In addition, the proportion of herding differences is also high (68%~85%), indicating that a simple herding behavior is not enough to cope with the complex investment environment currently; hence, the ROI worsens (dotted line in Figure 4). To briefly conclude, only with sufficient investment literacy and skills and by avoiding herding behavior, investment return can be assured and ultimately help households avoid investment predicaments.

Table 8. Decomposition results of differences in ROI.

Quantile	Difference in ROI	Characteristic Differences	Herding Differences	Ratios of Herding Effect
q10	−0.209	−0.036	−0.173	0.826
q20	−0.201	−0.054	−0.147	0.731
q30	−0.217	−0.068	−0.149	0.687
q40	−0.261	−0.076	−0.185	0.709
q50	−0.338	−0.056	−0.282	0.835
q60	−0.327	−0.048	−0.278	0.852
q70	−0.308	−0.059	−0.25	0.81
q80	−0.412	−0.081	−0.33	0.802
q90	−0.691	−0.134	−0.557	0.807

5. Discussion

5.1. Theoretical Implications

When investors follow the crowd instead of conducting their own analysis, herding behavior emerges in the financial sector, leading to unwarranted market rallies and sell-offs and triggering large-scale price fluctuations (Christie & Huang, 1995; Hwang & Salmon, 2006). Herding can be both rational and irrational. Rational investors gather information by observing the actions of other agents. In this study, we introduce a subjective, individual measurement method that allows investors to self-report their herding behavior. By doing so, we can understand whether herding behavior can be explained by certain investor attributes.

Our results confirm, at the individual level, previous findings that herding behavior is prevalent among individual investors in emerging markets such as China (Demirer & Kutun, 2006; Tan et al., 2008). Moreover, the use of quantile regression in this study reveals that the impact of herding varies across different return quantiles, which traditional mean-based models often overlook (Saastamoinen, 2008; Ma et al., 2018). This demonstrates the heterogeneity of behavioral effects and highlights the advantage of distribution-sensitive modeling in exploring psychological biases in investment behavior.

5.2. Practical Implications

In both the investment return and ROI models, herding behavior has a strong negative effect; the absence of herding behavior can reduce losses and is essential to increasing profits from stock investments. For investors seeking profit, avoiding herding behavior not only increases the investment return but also results in a higher ROI.

These findings highlight the importance of investor education programs and regulatory interventions aimed at curbing irrational crowd-following behaviors (Hilgert et al., 2003; van Rooij et al., 2011). Specifically, awareness campaigns and financial literacy initiatives targeting low-return households may enhance rational decision-making (Bernheim et al., 2001; Campbell, 2006). In light of the growing influence of social networks and digital finance on retail investor behavior in China (Wang et al., 2022), there is a pressing need for policies that counteract the amplification of herding through unverified or speculative information sources.

5.3. Research Limitations and Directions

While our analysis uses self-reported measures, which provide valuable subjective insights, these variables may be affected by cognitive biases or social desirability effects. This limitation has been discussed in behavioral literature, particularly regarding survey-based financial decision research (Cole et al., 2012; Hira & Loibl, 2005). The subjectivity inherent in the herding variable could lead to under- or over-estimation of actual behavioral patterns. Additionally, this study is based on cross-sectional data, which restricts our ability to track changes over time or in response to market events.

Future studies incorporating longitudinal panel data or experimental approaches may further refine our understanding of the temporal and causal dynamics behind household herding. Comparative studies across countries with different financial infrastructures and investor cultures may also shed light on how contextual factors shape the heterogeneity of herding behavior (E. C. Chang et al., 2000; Bashir et al., 2024).

6. Conclusions

The major finding from employing quantile regression-based decomposition is that regardless of whether the investment return or ROI is used as dependent variable, as the quantile increases, the differences in stock investment profitability and the herding effect

expand in opposite directions. Moreover, the herding effect's high proportion in total differences provides solid evidence of it being the main reason for differences in stock investment profitability. The quantile regression decomposition reinforces our finding that differences in the herding effect are a major factor influencing investment returns. Furthermore, the empirical results presented in this study provide support for the four hypotheses proposed in Section 2.6. First, the significant and negative association between self-reported herding behavior and both investment return and ROI confirms Hypothesis 1, indicating that households exhibiting stronger herding tendencies tend to experience poorer investment outcomes. Second, the quantile regression results reveal that the impact of herding is not uniform across the distribution of returns. The negative effect becomes more pronounced at higher quantiles, particularly beyond q75, thus supporting Hypothesis 2 and underscoring the importance of modeling return heterogeneity. Third, the interaction effects between herding behavior and various household characteristics—such as age, gender, income, and asset levels—demonstrate significant moderation effects. For example, income and gender show varying effects across quantiles in both investment and ROI models. These findings lend support to Hypothesis 3, which posits that demographic and financial characteristics condition the effect of herding. Finally, our analysis confirms Hypothesis 4, with the perceived credibility of market information moderating the negative association between herding and investment outcomes. Households that report higher trust in disclosed market information tend to suffer less from herding-induced losses or achieve better returns, suggesting that credible information can mitigate behavioral biases. The herding effect is detrimental to investment. Nevertheless, investors with adequate investment literacy and skills, along with the adoption of other rational methods, can avoid the pitfalls of herding behavior.

Regarding the total investment return as the dependent variable, the age variable shows a positive impact on investment returns after q90, indicating that older age positively influences investment effectiveness at higher quantiles. The cash holdings exhibit inconsistent effects across quantiles, suggesting that their influence on investment returns may depend heavily on underlying motivations, such as precautionary savings or market timing. Assets show significant effects on investment returns across all quantiles except for q50. Income exhibits a steep positive effect on investment effectiveness above q75. The herding effect for the head of the household shows significantly negative coefficients across all quantiles except for q25. In terms of ROI, male heads of the household experience fewer losses than female heads. The age variable shows better ROI at q50 and q90. Unlike the previous results concerning investment returns, assets do not have a significant effect on ROI. However, income demonstrates a significant positive effect on ROI. The herding effect in the stock market also shows a significantly negative correlation with ROI, consistent with the previous results on investment returns. If the head of the household believes in the information disclosed by the market, it is more likely that investment losses are reduced or that investment returns increase. By employing self-reported data regarding herding behavior, interesting and valuable outcomes were obtained with the quantile regression estimation method. In terms of financial literacy education, the government could launch comprehensive investor education programs aimed at novice investors, including free public courses and official knowledge platforms. Events such as 'Rational Investment Month' and 'Anti-Speculation Week' can heighten awareness of market sentiment risks. Establishing financial content disclosure guidelines for self-media, requiring clear information sourcing and risk disclaimers, could enhance information quality. Collaborations with social platforms to introduce content review systems or delayed posting mechanisms for speculative posts, along with regularly published 'Hot Stock Risk Reports,' can collectively help mitigate excessive market herding behavior.

As the quantile estimation results vary significantly across distributions, it is imprudent to conclude that the observed differences in the data are solely due to this factor. However, similar results to those in the literature have been obtained. Compared with traditional methods such as the mean reversion process or the conditional distribution assumption, employing the quantile regression model provides a new perspective for relevant studies. Empirical evidence from markets such as the U.S., Hong Kong, and Japan shows increasing return dispersions, implying a weaker presence of herding behavior—consistent with the findings of Christie and Huang (Christie & Huang, 1995). In contrast, emerging markets like South Korea and Taiwan exhibit lower-equity-return dispersions, indicating stronger herding tendencies, which aligns with our findings (E. C. Chang et al., 2000). Furthermore, studies have identified cross-market herding effects in the U.S., UK, and Germany, where investor sentiment in one market can influence herding behavior in another, particularly among European markets (Economou et al., 2018). In fact, the different empirical results found in this study are inconsistent with past findings, obtained by focusing only on the mean. Therefore, in future research, combining theories and experimental evidence from economics and other social sciences can enhance our understanding of how rationality and emotions interact to create herding effects in the financial sector. By collecting more independent variables, insights regarding how investors' financial literacy, information asymmetry, and various investment portfolios influence their responses can be discovered. This behavior will reflect the response of a well-functioning market to public information. While several variables show statistically significant coefficients at specific quantiles, including risk literacy, social trust, and happiness, these results often lie near conventional significance thresholds or lack theoretical consensus. Accordingly, they should be interpreted with caution, and future research may help validate or clarify these relationships. Additionally, while some behavioral variables such as risk literacy show statistically significant effects at lower quantiles, these results remain close to the margin of significance. Hence, interpretations must be made with appropriate caution, and further empirical validation is warranted.

Author Contributions: Conceptualization, Y.-C.T.; Methodology, Y.-C.T.; Software, Y.-C.T.; Formal Analysis, Y.-C.T.; Resources, G.-C.W.; Data Curation, G.-C.W.; Writing—Original Draft Preparation, I.-F.H.; Writing—Review and Editing, I.-F.H.; Visualization, I.-F.H.; Supervision, Y.-C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of the data, which were obtained from Survey and Research Centre for China Household Finance, available at <https://chfs.swufe.edu.cn/datacenter/apply.html> (accessed on 7 June 2023), with the permission of Survey and Research Centre for China Household Finance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Barnes, M. L., & Hughes, A. W. (2002). *A quantile regression analysis of the cross section of stock market returns*. Capital Markets: Asset Pricing & Valuation. [CrossRef]
- Bashir, U., Kayani, U. N., Khan, S., Polat, A., Hussain, M., & Aysan, A. F. (2024). Investor sentiment and stock price crash risk: The mediating role of analyst herding. *Computers in Human Behavior Reports*, 13, 100371. [CrossRef]
- Becker, T. A., & Shabani, R. (2010). Outstanding debt and the household portfolio. *The Review of Financial Studies*, 23(7), 2900–2934. [CrossRef]
- Bernheim, B., Garrett, D. M., & Maki, D. M. (2001). Education and saving: The long-term effects of high school financial curriculum mandates. *Journal of Public Economics*, 80(3), 435–465. [CrossRef]
- Bhandari, G., & Deaves, R. (2006). The demographics of overconfidence. *Journal of Behavioral Finance*, 7(1), 5–11. [CrossRef]
- Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets. *IMF Staff Papers*, 47(3), 279–310. [CrossRef]

- Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, 61(4), 1553–1604. [CrossRef]
- Chang, C., & Lin, S. (2015). The effects of national culture and behavioral pitfalls on investors' decision-making: Herding behavior in international stock markets. *International Review of Economics & Finance*, 37, 380–392. [CrossRef]
- Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, 24(10), 1651–1679. [CrossRef]
- Chen, X., & Song, J. (2021). Influence path analysis of rural household portfolio selection: A empirical study using structural equation modelling method. *The Journal of Real Estate Finance and Economics*, 64(2), 298–322. [CrossRef]
- Chiang, T. C., & Li, J. (2012). Stock returns and risk: Evidence from quantile. *Journal of Risk and Financial Management*, 5(1), 20–58. [CrossRef]
- Christie, W. G., & Huang, R. D. (1995). Following the pied piper: Do individual returns herd around the market? *CFA Magazine*, 51(4), 31–37. [CrossRef]
- Chuang, C. C., & Kuan, C. M. (2005). A quantile regression analysis of return-volume relation: Evidence from the Taiwan and U.S. Institute of Economics. *Academia Economic Papers*, 33, 379–404. [CrossRef]
- Cole, S., Paulson, A., & Shastry, G. (2012). *Smart money: The effect of education on financial behavior* (Harvard Business School Finance Working Paper No. 09-071). Harvard Business School. [CrossRef]
- Demirer, R., & Kutun, A. M. (2006). Does herding behavior exist in Chinese stock markets? *Journal of International Financial Markets, Institutions and Money*, 16(2), 123–142. [CrossRef]
- Economou, F., Hassapis, C., & Philippas, N. (2018). Investors' fear and herding in the stock market. *Applied Economics*, 50(34–35), 3654–3663. [CrossRef]
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427–465. [CrossRef]
- Guney, Y., Kallinterakis, V., & Komba, G. (2017). Herding in frontier markets: Evidence from African stock exchanges. *Journal of International Financial Markets, Institutions and Money*, 47, 152–175. [CrossRef]
- Hilgert, M. A., Hogarth, J. M., & Beverly, S. G. (2003). Household financial management: The connection between knowledge and behavior. *Federal Reserve Bulletin*, 89, 309–322.
- Hira, T. K., & Loibl, C. (2005). Understanding the impact of employer-provided financial education on workplace satisfaction. *Journal of Consumer Affairs*, 39(1), 173–194. [CrossRef]
- Hwang, S., & Salmon, M. (2006). *Sentiment and beta herding*. Available online: <https://warwick.ac.uk/fac/soc/wbs/subjects/finance/events/recentevents/pastevents/behaviouralfinance1day/hwang.salmon06.pdf> (accessed on 7 June 2023).
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33. [CrossRef]
- Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156. [CrossRef]
- Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1), 23–43. [CrossRef]
- Li, L. H., Lee, S. S., & Luo, D. (2021). *Gender differences in stock market participation: Evidence from Chinese households*. WGSRN: Gender & Globalization (Topic). [CrossRef]
- Li, W., Song, C., Xu, S., & Yi, J. (2020). High sex ratios and household portfolio choice in China. *The Journal of Human Resources*, 57(2), 465–490. [CrossRef]
- Lin, H. W. (2011). Elucidating rational investment decisions and behavioral biases: Evidence from the Taiwanese stock market. *African Journal of Business Management*, 5, 1630–1641. Available online: https://academicjournals.org/article/article1380559091_Lin.pdf (accessed on 15 March 2023).
- Liu, Y., & Zhang, M. (2021). Is household registration system responsible for the limited participation of stock market in China? *Review of Behavioral Finance*, 13(3), 332–350. [CrossRef]
- Ma, C., Xiao, S., & Ma, Z. (2018). Investor sentiment and the prediction of stock returns: A quantile regression approach. *Applied Economics*, 50(50), 5401–5415. [CrossRef]
- Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4), 577–590. [CrossRef]
- Melly, B. (2006). *Estimation of counterfactual distributions using quantile regression* (Discussion Paper). SIAW. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=85776810e1ef7f88cf10c59817b7a41ae3e3d863> (accessed on 15 March 2023).
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693. [CrossRef]
- Pavlidis, E. G., & Vasilopoulos, K. (2020). Speculative bubbles in segmented markets: Evidence from Chinese cross-listed stocks. *Journal of International Money and Finance*, 109, 102222. [CrossRef]
- Saastamoinen, J. (2008). Quantile regression analysis of dispersion of stock returns: Evidence of herding? *Keskustelualoitteita*, 57, 1–19. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a25263883caf467c21c1fc00b4076c24ac0a3e08> (accessed on 15 March 2023).
- Sabir, S. A., Javed, T., Hameed, W. U., & Yousaf, H. Q. (2020). Demographic Factors and herding behaviour of investors: Moderating role of islamic religiosity. *Review of Applied Management and Social Sciences*, 3(2), 193–203. [CrossRef]

- Sajid, A. (2015). Factors affecting investment decision making: Evidence from equity fund managers and individual investors in Pakistan. *Research Journal of Finance and Accounting*, 6, 135–141.
- Talpsepp, T., & Tänav, A. (2021). Do gender, age and education affect herding in the real estate market? *Journal of Behavioral and Experimental Finance*, 32, 100571. [[CrossRef](#)]
- Tan, L., Chiang, T. C., Mason, J. R., & Nelling, E. (2008). Herding behavior in Chinese stock markets: An examination of A and B shares. *Pacific-Basin Finance Journal*, 16(1–2), 61–77. [[CrossRef](#)]
- van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2), 449–472. [[CrossRef](#)]
- Wang, J., Zhang, D., & Wang, Z. (2022). Digital finance, stock market participation and asset allocation of Chinese households. *Applied Economics Letters*, 30(14), 1870–1873. [[CrossRef](#)]
- Wermers, R. (1999). Mutual fund herding and the impact on stock prices. *The Journal of Finance*, 54(2), 581–622. [[CrossRef](#)]
- Wu, X., & Zhao, J. (2019). Risk sharing, siblings, and household equity investment: Evidence from urban China. *Journal of Population Economics*, 33(2), 461–482. [[CrossRef](#)]
- Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19, 801–817.
- Yao, J., Ma, C., & He, W. P. (2014). Investor herding behaviour of Chinese stock market. *International Review of Economics & Finance*, 29, 12–29. [[CrossRef](#)]
- Zhang, Y., Jia, Q., & Chen, C. (2021). Risk attitude, financial literacy and household consumption: Evidence from stock market crash in China. *Economic Modelling*, 94, 995–1006. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.