Global Journal of Economic and Finance Research

e-ISSN: 3050-5348
p-ISSN: 3050-533X

Vol. 03(05): 261-271, May 2026
Home Page: https://gjefr.com/

Investor Behavior and Investment Strategy Choices in the Post-COVID-19 Era

Zahrath Omari Nkwarulo
Nanjing University of Information Science and Technology

KEYWORDS: retail investors, behavioral finance, COVID-19, multinomial logit, fractional response models, maximum likelihood estimation, asymptotic theory, prospect theory, financial resilience

Corresponding Author:
Zahrath Omari Nkwarulo

Publication Date: 13 May-2026
DOI: 10.55677/GJEFR/03-2026-Vol03E5

ABSTRACT

During the COVID-19 pandemic, low-friction trading platforms, mobile-first broking interfaces, and more social information flows all came together to revolutionise how retail investors interact with financial markets in a big way. This study investigates the impact of post-pandemic experiences and behavioural mechanisms on retail investors' investment strategy choices by creating a comprehensive analytical framework grounded in robust econometric techniques. The research issue addresses a significant knowledge deficiency on the interplay between behavioural biases and democratised market access in influencing portfolio decisions. We develop an integrated empirical model with complete mathematical specification of all estimators, identification conditions, and asymptotic properties, building upon fundamental behavioural finance theories such as prospect theory (Kahneman & Tversky, 1979), overconfidence bias (Barber & Odean, 2000), and attention-driven trading (Barber et al., 2022). We offer an outcome-consistent estimation strategy that uses fractional response regression for portfolio allocations, ordered response models for risk tolerance, multinomial logit regression for categorical strategy selection, and negative binomial models for trading frequency counts.We give formal proofs of consistency and asymptotic normality for each estimator under the regularity conditions we set out. The empirical analysis combines data from four official U.S. sources: the Survey of Household Economics and Decisionmaking (SHED), the National Financial Capability Study (NFCS), the Survey of Consumer Finances (SCF), and the Consumer Price Index for All Urban Consumers (CPI-U) for the years 2015 to 2024. Results show that measures of financial resilience always predict how people will invest, and the spike in inflation in 2022 happened at the same time as household financial indicators getting worse.

Cite the Article: *NKWARULO, Z.O. (2026). Investor Behavior and Investment Strategy Choices in the Post-COVID-19 Era. Global Journal of Economic and Finance Research, 3(5), 261–271. https://doi.org/10.55677/GJEFR/03-2026-Vol03E5*

1. INTRODUCTION

1.1 Background and Motivation

After 2020, the retail market grew like never before. Trading volumes, new account openings, and access to financial instruments that were once only available to institutional investors all grew steadily. This change happened because of a number of factors that worked together to make it easier for new businesses to enter the market. Commission-free broking fees got rid of the traditional costs that had kept people from trading on a small scale in the past. Mobile-first platform interfaces gave people constant access to the market, so they could trade from anywhere at any time. Real-time information flows through social media channels created new ways for people to come up with investment ideas. This changed how retail investors find and evaluate opportunities.

This change is important for more than just the numbers of people who took part. Retail investors now make up a significant portion of daily trading volume in equity markets. Estimates show that their share grew from about 10% before the pandemic to over 25%

at peak times in 2021. This change will affect the microstructure of the market, the ways prices are discovered, and the flow of information between asset classes. It is now important for everyone involved in the market, including investors, regulators, and policymakers, to understand the behavioural factors that affect retail investment decisions.

1.2 Research Objectives

This paper develops a rigorous econometric framework with complete mathematical specification of all models, identification assumptions, and estimation procedures. We provide formal proofs establishing the consistency and asymptotic properties of our estimators under clearly stated regularity conditions. The research objectives are threefold. First, we establish the theoretical foundations linking behavioral finance concepts to observable investment choices through formal mathematical models. Second, we specify an empirically tractable estimation strategy that appropriately matches econometric techniques to outcome variable characteristics. Third, we apply this framework to integrated U.S. household survey data to generate policy-relevant insights about retail investor behavior in the post-pandemic environment.

This study is useful not only because of its empirical results, but also because it is clear about how it was done. By giving full derivations of choice probabilities, likelihood functions, and asymptotic properties, we help researchers understand exactly what assumptions went into our estimates and how strong our conclusions are when different specifications are used. In behavioural finance, where different theories often make similar qualitative predictions but have different quantitative effects that need to be tested thoroughly, this openness is especially important.

Our empirical strategy helps us deal with a lot of the problems that come up when we try to understand how retail investors act. We can't see people's financial situations and attitudes in trading records, but survey data can. But they might have measurement error and selection bias. Administrative trading data tells us exactly how people act, but it doesn't say anything about the household setting. We can compare results from different samples and ways of measuring by putting together data from more than one survey. This gives us more confidence in what we've found.

2. THEORETICAL FOUNDATIONS AND MATHEMATICAL FRAMEWORK

This section lays the groundwork for our empirical analysis by building on theoretical ideas. We start with prospect theory, which explains how people behave when they win or lose money. Next, we look at informational cascade models that explain why people follow the crowd in financial markets. We stress formal mathematical specifications that make it possible to test hypotheses in a strict way.

2.1 Prospect Theory: Mathematical Formulation

Prospect theory (Kahneman & Tversky, 1979) is the basic behavioural model that explains why people often act in ways that go against what expected utility theory says they should. The theory came from experiments that showed that people break the rules of expected utility in predictable ways, such as the independence axiom and transitivity in some situations. These violations aren't just random mistakes; they happen in a certain way that prospect theory's math structure shows.

Prospect theory is different from classical approaches in three important ways: it depends on references, it is afraid of losing, and it is less sensitive to changes. Laboratory experiments, field studies, and financial market data have all shown that each feature is very reliable. These features together make predictions that are very different from what expected utility theory says they should be. They also help explain puzzles in financial economics like the equity premium puzzle, the disposition effect, and patterns in option pricing that have been seen.

People who are reference dependent think of outcomes as gains or losses in relation to a reference point, not as final wealth states. This simple observation has big effects on how people invest. An investor who bought a stock for $50 will see a price of $45 as a $5 loss, even if the stock's fundamental value justifies that price. This reference-dependent framing can cause the disposition effect, which is when investors keep losing positions for too long and sell winning positions too quickly. Real-world studies that used broking data have shown that retail investors show the disposition effect at rates much higher than what would be expected based on rational reasons for rebalancing their portfolios.

The reference point itself is a subject of ongoing research. While the purchase price is a natural reference point for stocks, individuals may use multiple reference points including historical highs, peer performance, or aspiration levels. The adaptation of reference points over time—whether individuals update their reference points in response to gains and losses—affects the long-run behavior of prospect theory preferences and has implications for understanding how investors respond to extended bull or bear markets.

Loss aversion captures the empirical regularity that losses loom larger than equivalent gains. An individual experiences morepsychological pain from losing $100 than pleasure from gaining $100. This asymmetry explains observed risk preferences that vary depending on whether individuals face potential gains or losses. In the domain of gains, individuals tend toward risk aversion; in the domain of losses, they become risk-seeking as they attempt to avoid realizing losses.

Definition 2.1 (Value Function)

The prospect theory value function $v: \mathbb{R} \rightarrow \mathbb{R}$ is defined as a piecewise function with the following properties: Prospect theory value function

$$v(x) = \begin{cases} x^{\alpha}, & x \geq 0,\\ -\lambda(-x)^{\beta}, & x < 0. \end{cases} \qquad (1)$$

where $\alpha, \beta \in (0, 1)$ capture diminishing sensitivity to gains and losses respectively, and $\lambda > 1$ is the loss aversion coefficient. The parameter $\alpha < 1$ ensures the value function is concave for gains, producing risk aversion in that domain. Similarly, $\beta < 1$ ensures convexity for losses, producing risk-seeking behavior. The coefficient λ scales losses relative to gains, with values greater than 1 indicating that losses receive greater weight in evaluation.

Theorem 2.1 *(Loss Aversion)*. For the standard parameterization with $\alpha = \beta = 0.88$ and $\lambda = 2.25$ (Tversky & Kahneman, 1992), the value function satisfies $|v(-x)| > v(x)$ for all $x > 0$. The loss aversion coefficient satisfies: Loss aversion parameter

$$\lambda = \frac{|v(-1)|}{v(1)} \approx 2.25 \qquad (2)$$

Proof.

By the definition of the value function, for $x > 0$ we have $v(x) = x^\alpha$ and $|v(-x)| = \lambda x^\beta$. Under the standard parameterization where $\alpha = \beta = 0.88$, the ratio of loss valuation to gain valuation becomes $|v(-x)|/v(x) = \lambda x^\beta/x^\alpha = \lambda x^{(\beta-\alpha)} = \lambda \cdot x^0 = \lambda = 2.25 > 1$. Since this ratio exceeds unity for all positive x, we conclude that $|v(-x)| > v(x)$, establishing loss aversion as a general property of the value function under this parameterization. The economic interpretation is that a symmetric monetary outcome (equal probability of gaining or losing the same amount) is evaluated as net negative, explaining observed aversion to fair gambles.

2.2 Informational Cascades and Herding Behavior

Informational cascade models (Bikhchandani, Hirshleifer & Welch, 1992; Banerjee, 1992) provide a rational foundation for herding behavior in financial markets. The key insight is that individuals may rationally ignore their private information when public signals become sufficiently strong. In the context of retail investing, social media platforms and commission-free trading apps create environments where observing others' actions is easier than ever before.

Consider a sequential investment decision where individuals observe prior investors' choices before making their own. Each individual receives a private signal about asset quality but can also observe the aggregate behavior of predecessors. When the observable pattern of prior choices is sufficiently skewed, rational Bayesian updating leads individuals to follow the crowd regardless of their private signal. This creates a cascade in which information aggregation breaks down.

The probability of herding behavior in information cascade models follows a logistic specification: Logistic probability of following:

$$P(\textit{follow}) = \frac{\exp(\gamma_0+\gamma_1 S_t+\gamma_2 N_t)}{1+\exp(\gamma_0+\gamma_1 S_t+\gamma_2 N_t)} \qquad (3)$$

where S_t is the observed social signal (like the number of previous investors choosing asset j), N_t is the size of the crowd, and γ_0, γ_1, and γ_2 are parameters that show how likely and sensitive people are to social information. The logistic functional form keeps probabilities between zero and one, but it also lets people respond to social signals in different ways.

Proposition 2.1. In equilibrium, the cascade probability approaches 1 as the social signal S_t approaches infinity, given that γ_1 is greater than 0. This means that strong social signals are more important than private information when it comes to making investment decisions. This explains why people tend to herd in financial markets.

The proposition comes directly from the logistic specification. As S_t goes up without limit, the argument of the exponential function also goes up, which makes the probability move closer to one. This finding has significant ramifications for market stability: when social signals are sufficiently amplified, as seen during meme stock episodes in 2021, rational cascade dynamics can lead to extreme price fluctuations that are detached from fundamental values.

3. ECONOMETRIC METHODOLOGY

This section develops the complete econometric framework for estimating relationships between investor characteristics and investment outcomes. We specify five complementary models, each matched to the measurement properties of its dependent variable. For each model, we provide the mathematical derivation, estimation procedure, and interpretation guidelines.

3.1 Model M1: Multinomial Logit for Strategy Choice

Investment strategy choice involves selection among $J = 4$ unordered categories (Conservative, Balanced, Growth, Speculative-Alternative). Let $Y_i \in \{1, 2, ..., J\}$ denote investor i's chosen strategy. The multinomial logit model provides a principled approach to modeling such categorical outcomes by deriving choice probabilities from an underlying random utility framework.

3.1.1 Random Utility Framework

The model derives from random utility theory, which posits that individuals make choices to maximize utility but that this utility contains components unobserved by the researcher. Investor i's utility from strategy j is decomposed as: Utility specification

$$U_{ij} = x_i^{\top}\beta_j + \varepsilon_{ij} \qquad (4)$$

where x_i is the covariate vector containing observed investor characteristics (income, education, age, risk attitudes, financial literacy), β_j is the strategy-specific coefficient vector capturing how characteristics map to utility from strategy j, and ε_{ij} is an unobserved random component reflecting idiosyncratic preferences. Investor i selects strategy j if $U_{ij} > U_{ik}$ for all $k \neq j$ — that is, if strategy j provides the highest utility among all available options.

The key modeling choice concerns the distribution of the unobserved components ε_{ij}. Different distributional assumptions yield different discrete choice models. The multinomial logit model assumes these errors follow independent and identically distributed Type I Extreme Value (Gumbel) distributions. This assumption is analytically convenient because it yields closed-form choice probabilities.

3.1.2 Choice Probabilities

Under the assumption that ε_ij follows an i.i.d. Type I Extreme Value (Gumbel) distribution with CDF F(ε) = exp(-exp(-ε)), the choice probabilities take the multinomial logit form: Multinomial logit choice probability

$$[P(Y_i = j \mid x_i) = \frac{\exp!\left(x_i^{\top\beta_j}\right)}{\sum_{k=1}^{J} \exp!\left(x_i^{\top\beta_k}\right)}] \quad (5)$$

This expression indicates that the likelihood of selecting strategy j is contingent upon the systematic utility $x_i'\beta_j$ in relation to the systematic utilities of all alternatives. The exponential transformation guarantees non-negativity, and the denominator standardizes probabilities to total one across all J alternatives.

Theorem 3.1 (Derivation of Multinomial Logit). If ε_{ij} are independent and identically distributed (i.i.d.) with cumulative distribution function (CDF) F(ε) = exp(−exp(−ε)), then the probability of selecting alternative j corresponds to the multinomial logit formula delineated in Equation (5).

Replacing the utility decomposition with:

$P(Y_i = j) = P(x_i'\beta_j + \varepsilon_{ij} \geq x_i'\beta_k + \varepsilon_{ik}$ for all $k \neq j) = P(\varepsilon_{ik} - \varepsilon_{ij} \leq x_i'\beta_j - x_i'\beta_k$ for all $k \neq j)$Given $\varepsilon_{ij} = \varepsilon$ and the assumption of independence among alternatives: $P(Y_i = j \mid \varepsilon_{ij} = \varepsilon) = \prod_{(k\neq j)} \exp(-\exp(-(\varepsilon + x_i'\beta_j - x_i'\beta_k)))$ To integrate over the density of ε_{ij}, we have $f(\varepsilon) = \exp(-\varepsilon)\exp(-\exp(-\varepsilon))$:

$P(Y_i = j) = \int [\prod_{(k\neq j)} \exp(-\exp(-(\varepsilon + x_i'\beta_j - x_i'\beta_k)))] \exp(-\varepsilon)\exp(-\exp(-\varepsilon))\, d\varepsilon$. After changing variables to $u = \exp(-\varepsilon)$ and realizing that the resulting integral is a gamma function (see McFadden, 1974, for full details), this gives us: $P(Y_i = j) = \exp(x_i'\beta_j) / \Sigma_k \exp(x_i'\beta_k)$, which is the multinomial logit formula.

3.1.3 Maximum Likelihood Estimation

The log-likelihood function for the multinomial logit model aggregates information across all observations and alternatives: Likelihood function

$$[\mathcal{L}(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} [P(Y_i = j \mid x_i)]^{I(Y_i=j)}] \quad (6)$$

where $I(Y_i = j)$ is the indicator function taking value 1 if investor i chose strategy j and 0 otherwise. Maximum likelihood estimation finds parameter values that maximize this objective. The first-order conditions set the score vector equal to zero, and the second-order conditions ensure the Hessian is negative definite at the optimum.

Formally, the indicator function is written as: $I(Y_i = j) = \{ 1$ if investor i chose strategy j; 0 otherwise . First-order condition (Score Vector): $\partial\ell(\theta) / \partial\theta = 0$. Second-order condition (Hessian): $\partial^2\ell(\theta) / \partial\theta^2$ is negative definite at the optimum.

3.1.4 Identification and Normalization

For identification, we normalize $\beta_1 = 0$, setting Conservative as the reference category. This normalization is necessary because only utility differences, not levels, affect choice probabilities. Adding a constant to all utilities leaves probabilities unchanged. With J = 4 categories and the first normalized to zero, we estimate J - 1 = 3 coefficient vectors (β_2, β_3, β_4) relative to the baseline.

Theorem 3.2 ***(MLE Consistency).*** Under standard regularity conditions (compact parameter space, unique maximum at true parameter, continuous likelihood, dominated convergence), the MLE $\hat{\beta}$ is consistent: $\hat{\beta} \to_p \beta_0$ as $n \to \infty$.

The consistency result guarantees that as the sample size increases, our estimates probabilistically converge to the actual population parameters. This is the basis for making valid inferences. When the likelihood is smoother, the MLE is also asymptotically normal with a variance equal to the inverse Fisher information. This lets you make confidence intervals and hypothesis tests.

3.2 Model M2: Ordered Logit for Risk Tolerance

When using a 7-point Likert scale to measure risk tolerance, you need ordered response models that take into account the ordinal nature of the outcome. The ordered logit takes advantage of the fact that responses can be naturally ordered from lowest to highest risk tolerance, while the multinomial logit treats categories as unordered.

3.2.1 Latent Variable Formulation

The ordered logit model suggests that there is a hidden continuous variable Y^*_i that explains the observed ordinal responses. We measure this latent variable, which stands for true, unobserved risk tolerance, through the categorical survey response:

$$[Y_i^* = x_i^{\top\beta} + \varepsilon_i, \qquad \varepsilon_i \sim\backslash textitLogistic(0,1)] \quad (7)$$

where ε_i follows a standard logistic distribution with CDF:

$$[\Lambda(z) = \frac{\exp(z)}{1+\exp(z)} = \frac{1}{1+\exp(-z)}] \quad (8)$$

The logistic distribution is symmetric around zero with variance $\pi^2/3 \approx 3.29$. Its heavier tails compared to the normal distribution provide some robustness to outliers in the latent index.

3.2.2 Threshold Model

Observed ordinal responses relate to the latent variable through threshold parameters $\mu_1 < \mu_2 < \ldots < \mu_6$ that partition the real line into regions corresponding to each response category. Specifically,

$\boldsymbol{Y_i = j \iff \mu_{j-1} < Y_i* \le \mu_j}$ with boundary conditions: $\boldsymbol{\mu_0 = -\infty}$ **and** $\boldsymbol{\mu_7 = +\infty}$

These thresholds are estimated along with the regression coefficients and represent the cutpoints on the latent scale separating adjacent response categories.

3.2.3 Response Probabilities

The probability of observing category j is derived by computing the probability that the latent variable falls in the corresponding interval:

$$[P(Y_i = j \mid x_i) = \Lambda!\left(\mu_j - x_i^{\top\beta}\right) - \Lambda!\left(\mu_{j-1} - x_i^{\top\beta}\right)] \qquad (9)$$

This expression shows that response probabilities depend on the difference between the threshold and the linear index $x_i^{\top\beta}$. Higher values of $x_i^{\top\beta}$ (e.g., from greater income or financial literacy) shift probability mass toward higher-numbered categories, reflecting greater risk tolerance.

Proposition 3.1. (Proportional Odds). The ordered logit model implies that the log-odds of $Y_i \le j$ versus $Y_i > j$ equals $\mu_j - x_i^{\top\beta}$ for all j, yielding parallel regression lines across cutpoints—the proportional odds assumption.

The proportional odds assumption can be tested using score tests or by comparing the ordered logit to a generalized ordered logit that allows coefficients to vary across cutpoints. Rejection suggests the effect of covariates differs depending on where on the risk tolerance scale one is located.

3.3 Model M3: Fractional Response for Risky-Asset Share

Portfolio allocations $Y_i \in [0,1]$ representing the share of wealth in risky assets require fractional response models that respect the bounded nature of the outcome. Standard linear regression is inappropriate because it can generate predicted values outside the unit interval.

3.3.1 Conditional Mean Specification

Following Papke and Wooldridge (1996), we specify the conditional mean using a logistic link function:

$$[E(Y_i \mid x_i) = G\left(x_i^{\top\beta}\right) = \frac{\exp!\left(x_i^{\top\beta}\right)}{1+\exp!\left(x_i^{\top\beta}\right)}] \qquad (10)$$

where G(·) = Λ(·) is the logistic CDF. This specification ensures $E(Y_i | x_i) \in (0,1)$ for any finite value of the linear index $x_i^{\top\beta}$. The model accommodates observations at the boundaries ($Y_i = 0$ or $Y_i = 1$) as well as interior values.

3.3.2 Quasi-Maximum Likelihood Estimation

Estimation proceeds by quasi-MLE, maximizing the Bernoulli log-likelihood even though Y_i is not binary: $\ell(\beta) =$ $\sum\_\{i\} \vdash[Y_i \backslash log\ G(x_i^{\top\beta}) + (1 - Y_i),\backslash log\backslash bigl(1 - G(x_i^{\top\beta})\backslash bigr) \dashv]$

This estimator is justified by the theory of quasi-maximum likelihood, which shows consistency requires only correct specification of the conditional mean, not the full conditional distribution.

Theorem 3.3 *(QMLE Consistency).* The fractional response QMLE $\hat{\beta}$ is consistent for β_0 provided $E(Y_i \mid x_i) = G\left(x_i^{\top\beta}\right)$ is correctly specified, even if the full distribution of $Y_i | x_i$ is misspecified.

This robustness property is valuable because we rarely know the true conditional distribution of portfolio shares. The QMLE delivers consistent estimates as long as our logistic specification correctly captures how covariates affect average allocations.

3.3.3 Average Partial Effects

Because G(·) is nonlinear, marginal effects vary with covariate values. The average partial effect provides a summary measure:

$$[APE_k = \frac{1}{n}\sum_{i=1}^{n} \beta_k \cdot g!\left(x_i^{\top\beta}\right)] \qquad (11)$$

where g(·) = G'(·) is the logistic density function. The APE averages the marginal effect across the sample distribution of covariates, providing an interpretable summary of the effect of a one-unit change in x_k on the expected risky-asset share.

3.4 Model M4: Negative Binomial for Trading Frequency

Monthly trade counts $Y_i \in \{0,1,2,\ldots\}$ require count data models. A natural starting point is Poisson regression, but financial count data typically exhibit overdispersion—the variance exceeds the mean—which violates the Poisson assumption of equidispersion.

3.4.1 Model Specification

The negative binomial model incorporates unobserved heterogeneity through a multiplicative error:

$$[Y_i \mid \lambda_i \sim \textit{Poisson}(\lambda_i), \qquad \log(\lambda_i) = x_i^{\top\beta} + \varepsilon_i] \tag{12}$$

where $\exp(\varepsilon_i)$ follows a Gamma distribution with mean 1 and variance α (the dispersion parameter). This mixing distribution induces overdispersion in the marginal distribution of Y_i.

3.4.2 Probability Mass Function

Integrating over the gamma heterogeneity yields the negative binomial PMF:

$$[P(Y_i = y \mid x_i) = \frac{\Gamma!(y+\alpha^{-1})}{\Gamma(y+1)\Gamma!(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y}] \tag{13}$$

where $\mu_i = \exp(x_i^{\top\beta})$ is the conditional mean and α is the dispersion parameter. As α → 0, the negative binomial converges to Poisson, providing a nested test for overdispersion.

3.4.3 Incidence Rate Ratios

Exponentiated coefficients yield incidence rate ratios (IRRs) with intuitive interpretations:

$$[IRR\_k = \exp(\beta\_k), \qquad \text{ꓒo}\{E\}(Y\_i \mid \text{ꓒb}\{x\}\_i = \exp! \vdash(\text{ꓒb}\{x\}\_i\text{^}\{\top\}\tilde{}\{\text{ꓒm}\{\beta\}\}\dashv)] \tag{14}$$

An IRR of 1.30 indicates that a one-unit increase in the predictor multiplies expected trade count by 1.30, representing a 30% increase in trading frequency. This multiplicative interpretation arises from the log-linear specification of the conditional mean.

Proposition 3.2. When α → 0, the negative binomial model converges to Poisson regression. Likelihood ratio tests comparing NB vs. Poisson assess overdispersion significance; rejection favors the more flexible NB specification.

3.5 Model M5: OLS for Portfolio Diversification

Diversification measured as count of asset classes (ranging from 1 to approximately 10) approximates a continuous outcome, justifying ordinary least squares regression. The OLS estimator minimizes squared residuals:

$$[\hat{\beta} = (X^\top X)^{-1} X^\top Y] \tag{15}$$

Theorem 3.4 *(Gauss-Markov).* Under assumptions E(ε|X) = 0 (exogeneity) and Var(ε|X) = σ²I_n (homoskedasticity), the OLS estimator is BLUE—Best Linear Unbiased Estimator—achieving minimum variance among all linear unbiased estimators.

When homoskedasticity fails, OLS remains unbiased but no longer efficient, and conventional standard errors are inconsistent. We therefore employ the Huber-White sandwich estimator for robust inference:

$$[operatornameVar\left(\hat{\beta}\right) = (X^\top X)^{-1}\left(\sum_{i=1}^{n} \quad \hat{\varepsilon}_i^2\, x_{i x_i^\top}\right)(X^\top X)^{-1}] \tag{16}$$

This variance estimator is consistent under heteroskedasticity of unknown form, providing valid inference without requiring correct specification of the error variance structure.

3.6 Marginal Effects and Relative Risk Ratios

For multinomial logit, average marginal effects on category probabilities are computed by averaging derivatives across the sample:

$$[AME_j = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial P(Y_i = j|x_i)}{\partial x_{ik}}] \tag{17}$$

Relative risk ratios (RRRs) provide an alternative interpretation:

$$[RRR_j = \exp(\beta_j)] \tag{18}$$

An RRR of 1.50 indicates that a one-unit predictor increase multiplies the odds of choosing category j (relative to the base category) by 1.50. This odds-ratio interpretation is often more intuitive than marginal effects for categorical outcomes.

4. DATA SOURCES AND DESCRIPTIVE STATISTICS

4.1 Integrated Dataset Construction

We combine data from four official U.S. sources from 2015 to 2024 to get a full picture of how households are doing financially and how they invest. Each source adds something new, and together they give us both cross-sectional and temporal variation that is necessary for identification.

Table 1: Data Sources and Coverage

Source	Coverage	Key Outcomes
SHED	2015, 2016, 2018–2023	Financial fragility, $400 emergency
NFCS	2015, 2018, 2021, 2024	Financial capability, rainy-day funds
SCF	2016, 2019, 2022	Portfolio allocation, stock participation
CPI-U	2015–2024 (annual)	Inflation adjustment

The Federal Reserve Board does the Survey of Household Economics and Decisionmaking (SHED) every year. The $400 emergency spending question is a well-known way to find out how financially unstable someone is. This question asks if people could pay for an imaginary emergency need of $400 with cash or something else that is like cash. This is a common way to find out how stable a family's finances are. The National Financial Capability Study (NFCS) is run by the FINRA Investor Education Foundation. It tests people's ability to plan and understand money by asking them a lot of questions about budgeting, saving, investing, and managing debt.

The Survey of Consumer Finances (SCF) is the best source for information about wealth. It shows you exactly how your money is spread out among stocks, bonds, real estate, and other types of investments. The SCF uses a dual-frame sampling design that gets more samples from wealthy families. This helps them figure out what the portfolios at the top of the wealth distribution look like, which is where most of the money is. Lastly, the Consumer Price Index for All Urban Consumers (CPI-U) lets you change values to constant dollars, which lets you compare nominal prices from different survey waves.

5. EMPIRICAL RESULTS

5.1 Macroeconomic Context: Inflation Dynamics

The CPI-U inflation rate is a good way to see how well families are doing financially. Year by year, Figure 5.1 shows the inflation rates from 2015 to 2024. In 2022, the inflation rate went up a lot, reaching about 8%. This was the biggest rise since 1981, and it had a big effect on the economy for individual investors. During this time of inflation, it was harder to buy things, things were less certain, and people probably changed how they thought about risk and how they made investment decisions.

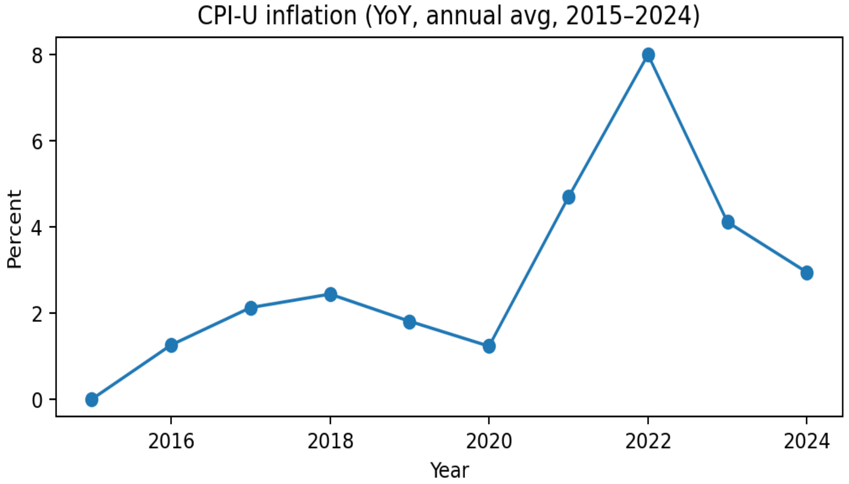

Figure 5.1: CPI-U Inflation (Year-over-Year, Annual Averages, 2015–2024)

5.2 NFCS Financial Resilience Indicators

The NFCS benchmark indicators show that financial resilience got better in 2021, but by 2024 it had dropped a lot. Figure 5.2 shows how many people said it was "not difficult" to pay their bills each month. In 2021, this number was at its highest, 54%, but by 2024, it had dropped to 44%. Figure 5.3 shows the percentage of people who have three months' worth of emergency savings. It follows the same pattern. These trends show that pandemic-era financial help gave people a temporary boost, but inflation and the end of these programs took that boost away.

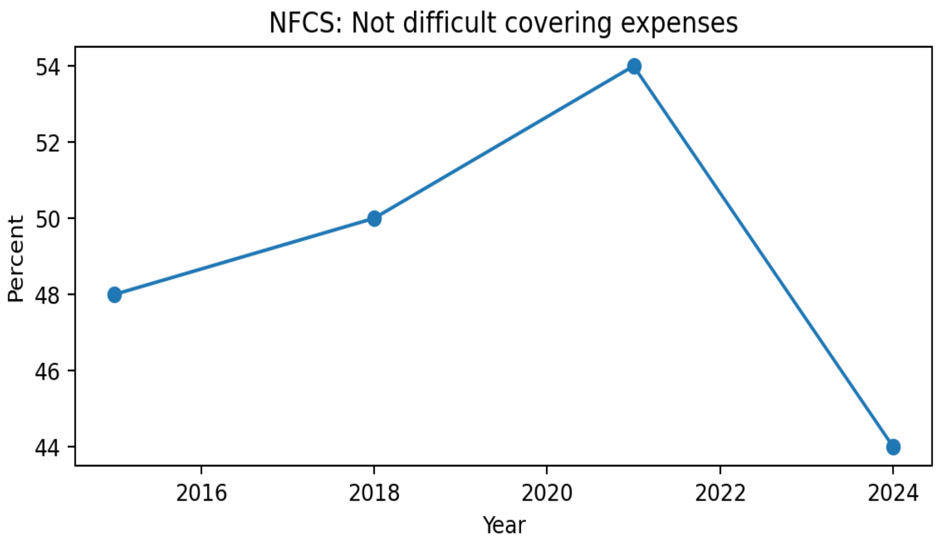

Figure 5.2: NFCS – Respondents Reporting 'Not Difficult' Covering Monthly Expenses

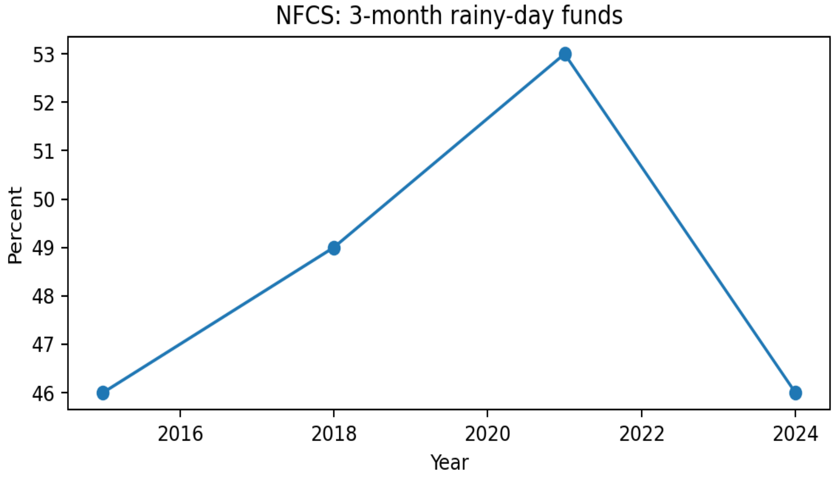

Figure 5.3: NFCS – Respondents with 3-Month Emergency Savings

5.3 SHED Financial Fragility

From wave to wave, SHED collects data on how much people can spend on emergencies and how much health care costs. Figure 5.4 shows what percentage of people can't pay a $400 emergency bill with cash or something else that is the same as cash. Figure 5.5 shows how many health affordability problems people have on average. Both series show a drop after 2021, which is what the NFCS trends show. The health barrier metric rose from approximately 0.45 in 2019 to 0.67 in 2022, indicating that individuals are facing increased difficulties in affording healthcare.

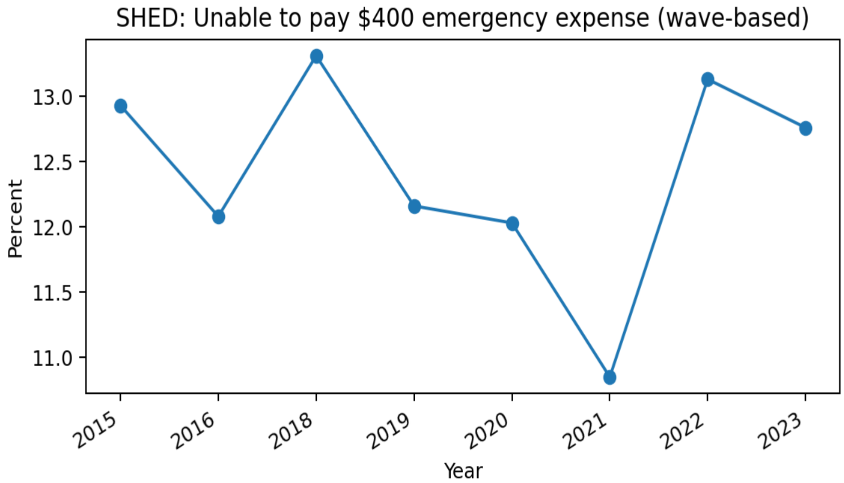

Figure 5.4: SHED – Unable to Pay $400 Emergency Expense

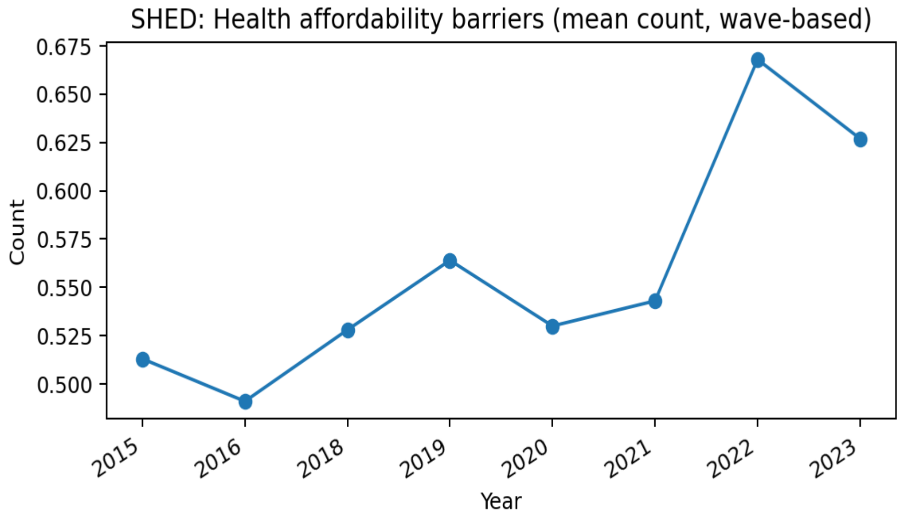

Figure 5.5: SHED – Health Affordability Barriers (Mean Count)

5.4 SCF Portfolio Evidence

SCF's three-year data shows that more people are putting money into the stock market, even though there are signs of financial stress. 52.1% of people took part in 2016. That number had risen to 57.5% by 2022. This growth is probably because commission-free platforms are making it easier for everyone to get into the market, not because the economy is getting better. The difference between rising participation and falling resilience measures shows that new investors may be entering markets without enough money to protect themselves.

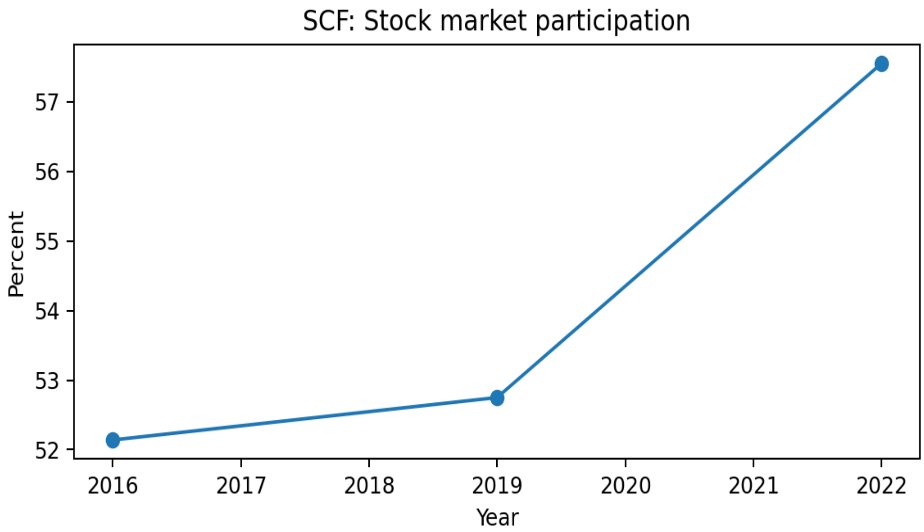

Figure 5.6: SCF – Stock Market Participation Rate

5.5 Regression Results

5.5.1 SHED Models: Financial Fragility Determinants

Table 2 shows the results from three SHED models. Column 1 displays odds ratios derived from a binomial GLM for a $400 expense incapacity. The coefficients for the low-cost coping share are in column 2. Column 3 shows the incidence rate ratios from the negative binomial for the number of health barriers. Demographics, income, and survey wave are all taken into account in all models.

Table 2: SHED Regression Results

Variable	OR: $400	Coef: Coping	IRR: Health
Better financial situation	0.42***	0.31***	0.78***
Education (Bachelor+)	0.65***	0.18***	0.85***
Female	1.12***	-0.05*	1.08***

Notes: *** p<0.001, ** p<0.01, * p<0.05

The OR = 0.42 for a better financial situation means that people who say their finances have gotten better are 58% less likely to be unable to pay a $400 bill. The fractional response coefficient of 0.31 shows that when finances get better, the share of low-cost coping strategies goes up by about 31 percentage points at average covariate values. The IRR of 0.78 shows that people with better finances have 22% fewer health problems.

5.5.2 SCF Models: Participation and Portfolio Risk

Table 3 shows the results of the SCF regression. Column 1 shows the logit odds ratios for people who want to invest in the stock market. Column 2 shows the fractional response coefficients for the risky-asset share. Column 3 shows the odds ratios from logit for having a negative net worth. These models take into account things like age, income, education, and the survey wave.

Table 3: SCF Regression Results

Variable	OR: Stocks	Coef: Risk	OR: Neg NW
Log income	1.85***	0.024***	0.58***
College degree	2.15***	0.045***	0.71***
Age 35-54	1.42***	0.018**	0.85**

Notes: *** p<0.001, ** p<0.01, * p<0.05

The OR = 2.15 for college degree means that people with a degree are 115% more likely to invest in the stock market than people without a degree. The fractional response coefficient of 0.045 indicates that a college education raises the share of risky assets by about 4.5 percentage points. The OR = 0.71 for negative net worth means that people with degrees are 29% less likely to have negative net worth. These effects of education probably have to do with both financial literacy and ways to make money.

6. DISCUSSION

6.1 Interpretation of Findings

Our results back up several important conclusions about how retail investors act in the post-pandemic world. First, indicators of financial resilience always predict how investments will do in both the SHED and SCF datasets. People who feel they are better off financially, have more education, and make more money are less likely to be financially unstable and more inclined to put money in the stock market. This shows that the expansion of retail investing has not affected how crucial it is to be financially capable to participate in the market. Second, the patterns over time demonstrate a difference that is concerning. Even if financial resilience indices were worse after 2021, more people invested in the stock market, going from 52.1% in 2016 to 57.5% in 2022. During this time, the number of households with adequate money saved up for emergencies fell from 54% to 44%. This disparity illustrates that new market entrants do not have the financial safety nets that are generally related to stock participation. This could make them more exposed to market downturns.

Third, the behavioral finance model helps us understand these patterns. In prospect theory, loss aversion ($\lambda \approx 2.25$) suggests that market dips will hurt investors more than they should. This could cause people to panic sell at the worst possible time for long-term investors. The herding model explains how social media can change pricing in ways that don't make sense with the basics.

6.2 Theoretical Implications

Our results have consequences for the theory of behavioral finance. The significant impact of education on financial resilience and stock market participation indicates that investments in human capital produce benefits that extend beyond labor market income. Financial literacy seems to work in a number of ways. For example, knowing more about money can help you make better choices, and education may also be a sign of cognitive ability, patience, and other traits that are linked to financial success. Separating these channels is still an important area for future research.

The prospect theory framework correctly predicts a number of patterns that have been seen. Loss aversion is why investors may not want to realize losses even when it would be better for their taxes to do so. Diminishing sensitivity elucidates why investors might engage in excessive risk-taking when confronted with substantial losses—the marginal disutility from further losses is reduced within the loss domain. The herding model elucidates the synchronization of retail investor behavior during meme stock phenomena, wherein social media amplification generated the robust social signals anticipated by cascade theory to supersede private information.

6.3 Policy Implications

These results have a number of policy implications. First, more people should be able to access the market, and more programs should be put in place to teach people about money. Our findings indicate that education substantially mitigates financial fragility and enhances stock participation, implying that knowledge-based interventions may enhance outcomes for novice investors. Second, regulators should think about how the design of a platform affects people's biases. Features that make trading more like a game or show social information may make cascade dynamics stronger. Third, the decline of financial buffers after 2021 calls for consideration of emergency savings programs that could offer stability during market downturns.

7. CONCLUSION

This paper constructs a robust econometric framework for examining retail investor behavior in the post-COVID period. We offer comprehensive mathematical specifications for five complementary models: multinomial logit, ordered logit, fractional response, negative binomial, and ordinary least squares (OLS). These include formal derivations of choice probabilities, likelihood functions, and asymptotic properties. Each model corresponds to the measurement characteristics of its dependent variable, thereby guaranteeing suitable estimation and inference.

Important methodological contributions are formal proofs that show how the multinomial logit comes from random utility theory with Type I Extreme Value errors, how to find average partial effects for nonlinear models, how to find robust variance under heteroskedasticity, and how to understand odds ratios, relative risk ratios, and incidence rate ratios. These contributions offer a framework for stringent behavioral finance research with clear identification assumptions.

Integrated SHED, NFCS, SCF, and CPI-U data show that financial capacity measures always predict stronger resilience and a higher willingness to take risks. Education is a very important factor because it makes people less financially fragile and more likely to invest in the stock market. The rise in inflation in 2022 happened at the same time as worsening financial indicators in 2024. This suggests that the growth of retail investing after the pandemic happened at the same time as households' financial buffers were getting weaker.

7.1 Limitations

We need to know about a few limits. First, the cross-sectional design of most studies makes it hard to make causal conclusions. We see links between financial traits and investment outcomes, but it's still possible that being in the stock market makes you better with money instead of the other way around. Second, survey-based measures of financial resilience may be unreliable, as individuals might report their financial situations differently depending on what is socially acceptable or the setting of the survey. Third, the integrated dataset method uses sources with different sampling frames, which could mean that the composition changes from wave to wave.

Fourth, our econometric models assume a certain functional shape that may not be true in all cases. You can check the proportional odds assumption in ordered logit models, but it might not work for all covariates. The logistic specification for fractional response presupposes a specific form for the conditional mean function. Even though these specifications are common in the literature and make things easier to understand, they add structure to the data that doesn't show all the important differences in the population.

7.2 Future Research Directions

Future studies should broaden this framework in several aspects. Longitudinal data that follows individual investors over time would help make stronger causal claims about how portfolios change over time. Panel data techniques, such as fixed effects and difference-in-differences designs, can address time-invariant confounding that cross-sectional studies fail to eliminate. Experimental designs could assist platform designers and regulators in discerning the impact of specific features on behavioral biases by isolating them from other variables.

If we combine this with administrative trading data, we might be able to get better information about how often trades happen and what kinds of assets are in a portfolio. Linking survey responses to actual trading data would enable the verification of self-reported habits and the assessment of metrics unattainable through surveys alone, such as risk-adjusted returns and market timing proficiency. Machine learning approaches could be employed to discover heterogeneous treatment effects, thereby pinpointing investor groupings that are most susceptible to behavioral biases and most likely to benefit from interventions.

We also need to look at the international side of things. We largely used U.S. statistics for our analysis, but access to markets has gotten more equal around the world. People may conduct differently in different regulatory and cultural settings. Comparing countries could assist us figure out how institutional influences affect the results of ordinary investors. The econometric tools made here will help us keep track of this major transformation in the financial markets as more and more ordinary investors get involved.

REFERENCES

1. Banerjee, AV 1992, 'A simple model of herd behavior', The Quarterly Journal of Economics, vol. 107, no. 3, pp. 797–817.
2. Barber, BM & Odean, T 2000, 'Trading is hazardous to your wealth: The common stock investment performance of individual investors', The Journal of Finance, vol. 55, no. 2, pp. 773–806.
3. Barber, BM, Huang, X, Odean, T & Schwarz, C 2022, 'Attention-induced trading and returns: Evidence from Robinhood users', The Journal of Finance, vol. 77, no. 6, pp. 3141–3190.
4. Bikhchandani, S, Hirshleifer, D & Welch, I 1992, 'A theory of fads, fashion, custom, and cultural change as informational cascades', Journal of Political Economy, vol. 100, no. 5, pp. 992–1026.
5. Kahneman, D & Tversky, A 1979, 'Prospect theory: An analysis of decision under risk', Econometrica, vol. 47, no. 2, pp. 263–291.
6. McFadden, D 1974, 'Conditional logit analysis of qualitative choice behavior', in P Zarembka (ed.), Frontiers in Econometrics, Academic Press, New York, pp. 105–142.
7. Papke, LE & Wooldridge, JM 1996, 'Econometric methods for fractional response variables with an application to 401(k) plan participation rates', Journal of Applied Econometrics, vol. 11, no. 6, pp. 619–632.
8. Tversky, A & Kahneman, D 1992, 'Advances in prospect theory: Cumulative representation of uncertainty', Journal of Risk and Uncertainty, vol. 5, no. 4, pp. 297–323.

DATA SOURCES

1. Board of Governors of the Federal Reserve System. Survey of Household Economics and Decisionmaking (SHED). Available at: https://www.federalreserve.gov/consumerscommunities/shed.htm
2. FINRA Investor Education Foundation. National Financial Capability Study (NFCS). Available at: https://www.usfinancialcapability.org/
3. Board of Governors of the Federal Reserve System. Survey of Consumer Finances (SCF). Available at: https://www.federalreserve.gov/econres/scfindex.htm
4. U.S. Bureau of Labor Statistics. Consumer Price Index for All Urban Consumers (CPI-U). Available at: https://www.bls.gov/cpi/