

Louvain School of Management

Herding Behavior in Individual Investors: Evidence & Market Consequences

Author: Yoan Bossart
Supervisor: Catherine D'Hondt
Academic year: 2025-2026
Master's in management sciences - Finance & Transition
Herding Behavior in Individual Investors: Evidence & Market Consequences
Daytime schedule

A. Declaration regarding AI use

<u>Statements on responsible use of generative AI</u>	<u>Your answer</u>
The content of my master's thesis is my original output, even if I used generative AI to help prepare it.	☒ Yes ☐ No
I/we master the theories, tools and methods that I use for the thesis.	☒ Yes ☐ No
I verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.	☒ Yes ☐ No
I acknowledge that I master the final output and that I am able to explain it and answer questions about it.	☒ Yes ☐ No
I made sure not to provide generative AI tools with access to confidential data or information	☒ Yes ☐ No

Signature

<u>Author last name, first name</u>	<u>Date</u>	<u>Signature</u>
1 Bossart, Yoan	May 26 th , 2026	<i>Yoan Bossart</i>

Note

In the process of writing this master's thesis, the author utilized the following AI tools for limited support purposes:

1. *DeepL AI*: occasionally used to assist with the translation of specific technical terms from French to English, in order to ensure linguistic accuracy.
2. *ChatGPT AI*: used as a linguistic support tool to improve wording and refine the structure of certain passages.
3. *Claude AI*: used as a programming assistant to support the structuring and debugging of parts of the R code used in the empirical analysis.
4. *Perplexity AI*: used selectively to explore alternative coding approaches and identify potential optimizations.

The author has reviewed and edited all outputs provided by the use of these AI tools. The conceptual development, methodological choices, empirical analysis, and final writing of the thesis remain entirely the responsibility of the author. AI tools were used only for support tasks such as wording and coding assistance.

By signing this declaration, the author affirms that the content of this master's thesis reflects their original work, supported only by the responsible and limited use of AI tools.

B. Acknowledgements

I would first like to express my sincere gratitude to Professor Catherine D'Hondt for her trust and her assistance, as well as for giving access to the dataset. Moreover, thank you to the organizers of the *Rallye Boursier de L'Echo* for making this dataset available for academic purposes.

A special thank you to Professors Patrick Scarmure, Frédéric Vrins, and Mikael Petitjean for their teaching, advice, and encouragement. I would also like to thank the professors of finance at the Louvain School of Management for the quality of their teaching throughout my studies.

I am grateful to Mr. Roee Boosi for his openness and for sharing valuable insights on the GameStop event, which have been very useful for my work.

My deepest gratitude goes to my parents, grandparents, and Amélie Beauport for their patience, encouragement, and constant support throughout this journey.

I am also sincerely grateful to Odile Botman and Quentin Zèques for their kindness and assistance.

Finally, I would like to thank my friends for their support and encouragement, as well as for all the discussions and moments shared during this demanding but rewarding academic journey.

C. List of abbreviations

Abbreviation	Meaning
Adj. R^2	Adjusted R-squared
AF	Adjustment Factor
EMH	Efficient Market Hypothesis
ETF	Exchange-Traded Fund
ew	Equal-Weighted
FE	Fixed Effects
FOMO	Fear of Missing Out
GME	GameStop Corp. (ticker)
HA-HE	Hypotheses A through E
ID	Investor Identifier
IHM	Individual Herding Measure
$\overline{IHM}$	Average Individual Herding Measure
IHM^{ew}	Equal-weighted Individual Herding Measure
IHM^{vw}	Value-weighted Individual Herding Measure
ISIN	International Securities Identification Number
LSV	Lakonishok, Shleifer and Vishny (1992) herding measure (LSV)
OLS	Ordinary Least Squares
pp.	Percentage points
R^2	Coefficient of Determination
SD	Standard Deviation
SE	Standard Error
SLSV	Signed LSV (buy vs sell herding)
TTWO	Take-Two Interactive Software, Inc. (ticker)
vw	Value-Weighted
w R^2	Within R-squared

D. List of symbols and notation

Symbol	Meaning
Δ	Change or difference
i	Investor index
j	Stock index
t	Time index (month)
$E [\bullet]$	Expectation operator
ρ	Correlation coefficient
α	Intercept term
β	Regression coefficient
λ_t	Time fixed effects
$\varepsilon_{i,t}$	Error term for investor i at time t
$P_{j,t}$	Average price of stock j during month t (<i>IHM</i>)
p_t	Market-wide proportion of buyers (<i>LSV / SLSV</i>)
$p_{j,t}$	Proportion of buyers for stock j at time t (purchase intensity) (<i>LSV / SLSV</i>)
$I_{j,t}$	Number of active investors trading stock j at time t (<i>LSV</i>)
$B_{j,t}$	Number of net buyers for stock j at time t (<i>LSV</i>)
$S_{j,t}$	Number of net sellers for stock j at time t (<i>LSV</i>)
$AF_{j,t}$	Adjustment factor (<i>LSV</i>)
$X_{j,t}$	Number of net buyers (binomial variable) (<i>LSV</i>)
n	Number of trials (binomial parameter) (<i>LSV</i>)
N	Number of observations (sample)
$QuizScore_i$	Financial knowledge score of investor i
$QUIZ_OK_i \mid QUIZ_NOK_i$	Number of correct incorrect quiz answers for investor i
$IntlDiv_i$	International diversification of investor i
$\sim$	'Is distributed as' (distribution notation)
*	Statistically significant at the 10% level ($p < 0.1$)
**	Statistically significant at the 5% level ($p < 0.05$)
***	Statistically significant at the 1% level ($p < 0.01$)

E. Table of contents

A. Declaration regarding AI use.....	II
B. Acknowledgements.....	IV
C. List of abbreviations.....	V
D. List of symbols and notation	VI
E. Table of contents	VII
F. List of tables.....	X
F.1. Main text.....	X
F.2. Thematic appendices	X
F.3. Table appendix.....	X
G. List of figures	XI
1. Introduction.....	1
1.1. Context and relevance of the topic.....	1
1.2. Problem statement and structure of the thesis.....	2
2. Theoretical framework and literature review	3
2.1. Market efficiency and behavioral deviations	3
2.2. Emergence of behavioral finance.....	3
2.2.1. Expected Utility Theory & Prospect Theory.....	3
2.2.2. Mental accounting and framing effects	4
2.2.3. Cognitive and social drivers of investor behavior	4
2.3. Investor herding behavior: Definitions and mechanisms.....	4
2.4. Measurement of herding behavior	5
2.4.1. Asset-level herding: Lakonishok, Shleifer and Vishny measure (LSV)	5
2.4.2. Investor-level herding: the Individual Herding Measure.....	7
2.4.3. Alternative approaches and methodological debates	8
2.5. Determinants of herding behavior	8
2.5.1. Individual characteristics	8
2.5.2. Trading environment factors	9
2.5.3. Associated behavioral biases among ‘Rallye Boursier’ participants	9
2.5.4. Momentum versus contrarian tendencies.....	9
2.5.5. Sociodemographic and cultural factors	10
2.6. Digital catalysts of herding behavior.....	12
2.6.1. Attention-driven investing rise	12
2.6.2. Financial influencers (‘finfluencers’) role	12
3. Methodology.....	13
3.1. Research gaps and contribution.....	13
3.2. Research hypotheses	14
3.3. Data construction and cleaning procedures.....	15
3.3.1. Implementation overview (data pipeline).....	16
3.3.2. Sample selection and definition of active investors.....	16
3.3.3. Data cleaning and treatment of specific cases.....	16
3.3.4. Construction of the herding measure	17
3.3.5. Construction of explanatory variables	18
3.4. Empirical strategy.....	20
3.4.1. Panel data structure and sample definition	20
3.4.2. Model specification and key methodological choices.....	21
3.4.3. Hypothesis testing framework (Tests for each hypothesis)	22
3.4.4. Inference and model evaluation	24

3.5. Robustness strategy & checks	25
4. Results	26
4.1. Descriptive evidence on herding	26
4.1.1. Distribution of IHM (HA)	27
4.1.2. Stock-level herding (LSV Overview)	27
4.2. Determinants of herding (HB, HC, HD)	28
4.2.1. Baseline panel regressions (MIN_BUYS = 10)	28
4.2.2. Economic implications.....	29
4.2.3. Model fit and explanatory power	29
4.3. Herding and portfolio performance (HE)	30
4.4. Robustness analysis	32
4.4.1. Alternative activity thresholds	32
4.4.2. Alternative eligibility rule (MIN_TRADES)	32
4.4.3. Winsorization of performance	33
4.4.4. Equal-weighted IHM.....	33
4.4.5. Level vs Log specification of trading activity	33
4.4.6. LSV November sensitivity	33
5. Discussion.....	34
5.1. Interpretation of main findings	34
5.2. Contribution to the literature.....	35
5.3. Economic and market implications	36
6. Limitations and future research	37
7. Conclusion.....	39
8. Appendices.....	41
8.A. Appendix A. LSV adjustment factor computation	41
8.B. Appendix B. Equity universe filter	42
8.C. Appendix C. Numerical illustration of LSV, SLSV & IHM.....	43
8.C.1. Stock-level trading activity	43
8.C.2. Computation of LSV and SLSV	44
8.C.3. Investor-level herding (IHM)	45
8.C.4. Interpretation	46
8.D. Appendix D. Model fit interpretation and literature comparison.....	49
8.D.1. Why low R^2 is expected and acceptable	49
8.D.2. Comparison with Merli & Roger (2013).....	49
8.D.3. Within R^2 interpretation in FE models	50
8.E. Appendix E. Sensitivity to LSV computation	51
8.E.1. November 2019: Mechanical LSV equal to zero	51
8.E.2. Focus on December and January	52
8.E.3. Impact on final results.....	52
8.F. Appendix F. Sample construction & descriptive statistics	53
8.F.1. Sample construction & activity threshold.....	53
8.F.2. International diversification & geographical concentration.....	53
8.G. Appendix G. Additional empirical evidence	54
8.G.1. Coordination effects	54
8.G.2. Popularity and herding intensity (IHM)	54
8.G.3. Stock popularity and stock-level herding (LSV and SLSV)	56
8.G.4. Most herded stocks by month	57
8.H. Appendix H. Figures	58
8.H.1. IHM by month Boxplot q05 - q95	58

8.H.2. IHM Distribution (Histogram + Density): Full range	59
8.H.3. IHM Distribution (Histogram + Density): q01 - q99	60
8.H.4. Distribution of LSV (Stock-month).....	61
8.H.5. Distribution of Signed LSV (SLSV, stock-month)	62
8.H.6. IHM by language boxplot q05 - q95.....	63
8.H.7. Baseline determinants of herding	64
8.H.8. Investor participation and stock-level herding (LSV)	65
8.H.9. Investor participation and stock-level herding (LSV): Raw	66
8.H.10. Herding intensity and portfolio performance q01 - q99	67
8.H.11. Comparison between full (Nov - Jan) & restricted sample (Dec - Jan).....	68
8.I. Appendix I. Tables.....	69
8.I.1. Table I1. Winsorization robustness check (performance regression)....	69
8.I.2. Table I2. Stock-level herding statistics.....	70
8.I.3. Table I3. LSV stock-month diagnostics	71
8.I.4. Table I4. Value-weighted vs equal-weighted IHM	72
8.I.5. Table I5. Level vs Log specification	73
8.J. Appendix J. R Code	73
9. Bibliography	74

F. List of tables

F.1. Main text

Table 1. Descriptive statistics

Table 2. Baseline determinants of herding

Table 3. Herding and portfolio performance

F.2. Thematic appendices

Table C1. Trading activity by stock and month (*Example*)

Table C2. LSV and SLSV (*Example*)

Table C3. Investor trades (month t_2) (*Example*)

Table C4. $IHM_{i,t}^{ew}$ is not always positive (*Example*)

Table G2. Herding intensity and popular stocks

Table G3. Stock-level herding (LSV and SLSV) and popular stocks

Table G4. Top 3 buy-side & sell-side herding stocks by month

F.3. Table appendix

Table I1. Winsorization robustness check

Table I2. Stock-level herding statistics

Table I3. LSV stock-month diagnostics

Table I4. Value-weighted vs equal-weighted ($IHM_{i,t}^{vw}$ vs $IHM_{i,t}^{ew}$)

Table I5. Level vs log specification

G. List of figures

Figure H1. IHM by month Boxplot q05 - q95

Figure H2. IHM Distribution (Histogram + Density): Full range

Figure H3. IHM Distribution (Histogram + Density): q01 - q99

Figure H4. Distribution of LSV

Figure H5. Distribution of Signed LSV

Figure H6. IHM by language Boxplot q05 - q95

Figure H7. Baseline determinants of herding ($IHM_{i,t}^{vw}$)

Figure H8. Investor participation and stock-level herding (LSV)

Figure H9. Investor participation and stock-level herding (LSV) (Raw relationship)

Figure H10. Herding intensity and portfolio performance q01 - q99

Figure H11. Comparison between full (Nov - Jan) & restricted sample (Dec - Jan)

Note: All tables and figures are based on the author's calculations using the Rallye Boursier dataset.

1. Introduction

1.1. Context and relevance of the topic

Financial markets have changed substantially over the past two decades, notably through the democratization of investing access via online brokers. While markets were historically dominated by institutional investors, an increasing share of transactions is now executed by individuals. The growth of low-cost platforms such as Trade Republic, Degiro, and Robinhood has lowered entry barriers, allowing investors to invest even very small amounts, sometimes as little as one euro. At the same time, the digitalization of financial information through websites, forums, and social networks has increased the visibility of financial products, such as ETFs (Freibauer, Grawert, & Rieger, 2024).

The rise of retail investors now constitutes a force capable of influencing stock prices. The GameStop (GME) episode in 2021 is the most emblematic illustration. Through Reddit, hundreds of thousands of individual investors coordinated to buy GME shares, causing a historic short squeeze from hedge funds (Pomparelli, 2024) that pushed the stock from under \$5 (MarketWatch, 2025) to more than \$400 within weeks (Boosi, 2025). For many participants, the event also reflected community identity and a fervent opposition to Wall Street institutions associated with the 2008 crisis (Grant, 2021).

Other episodes point to similar dynamics, on a smaller scale. For instance, announcements related to GTA VI (TTWO) were rapidly priced in (Farhath, 2025) and widely amplified across social networks. This emphasizes the way information is diffused through online interactions (Sismanis, 2025). Comparable behaviors have been observed for NVIDIA (Investing.com, 2025) and Tesla stocks (Harring, 2023), as well as in crypto-asset markets including Bitcoin, Ethereum, and Dogecoin. These examples confirm that retail investors are now fully integrated into financial markets, that coordination can influence prices, and that their behavior cannot be fully explained by standard rational models (Wheat & Eckerd, 2025).

This thesis contributes to the literature in three main ways. First, it provides investor-level evidence on retail herding behavior using trading data from the *Rallye Boursier* competition. Second, it documents a positive relationship between investor sophistication and herding intensity, contrasting with the conventional expectation that more sophisticated investors behave more independently. Third, it examines the relationship between herding and raw portfolio performance, showing that stronger herding is associated with lower raw returns.

1.2. Problem statement and structure of the thesis

Considering the elements above, herding among retail investors appears to be an important subject of study for both academic research and market practitioners. The research question proposed is therefore:

How can retail herding be robustly measured and explained through the combined influence of socio-demographic, cultural, and market participation factors, and what implications does it have for portfolio performance?

To answer this question, this thesis is structured around three main axes:

- (i) Literature review:** discussion of the theoretical background, existing empirical work on herding, and recent studies on social networks, finfluencers, and sociodemographic differences.
- (ii) Empirical analysis:** study of herding on a sample of retail investors, using investor socio-demographic, order, and transaction data from the *Rallye Boursier*. This analysis notably considers age, sophistication, language, investor engagement, and performance.
- (iii) Discussion, conclusions and limitations:** interpretation of the results in relation to investor behavior, sophistication effects, popularity dynamics, and digital market environments.

This thesis aims to examine a phenomenon which connects financial markets, social psychology, and digital sociology. The objective is to contribute to a better understanding of the growing role of individual investors in today's markets in the era of social networks and 'finfluencers'. The following chapter reviews the main theoretical and empirical literature on herding among retail investors and the key concepts required to understand the topic.

2. Theoretical framework and literature review

2.1. Market efficiency and behavioral deviations

According to the Efficient Market Hypothesis (EMH), asset prices always reflect all available information. Systematic outperformance would therefore be impossible, as new information is assumed to be instantaneously incorporated into prices. Collective movements, often perceived as disorganized, should therefore not significantly affect asset valuation (Fama, 1970).

However, empirical research has progressively challenged this idealized view. Lakonishok, Shleifer and Vishny (1992) showed that even institutional investors are not always rational and can follow biased strategies. Later, Brown, Wei and Wermers (2007) provided additional evidence of peer imitation among fund managers, a line of research that has since been extended by Frey, Herbst, and Walter (2014). Biases such as overconfidence, search for social validation or even FOMO play a crucial role in herding (Broihanne & Capelle-Blancard, 2018). Indeed, after major financial crises, such as in 2008 (subprimes) and 2020 (Covid-19), herding has become increasingly observed. In fact, retail investors may replicate strategies of ‘perceived experts’ to reduce uncertainty, delegating their judgment. Thus, fear leads not only to panic selling but also to stronger herding (Banerjee, 1992). In this context, social dynamics may dissociate asset prices with fundamental values.

2.2. Emergence of behavioral finance

2.2.1. Expected Utility Theory & Prospect Theory

The foundations of decision theory come from Pascal (1670) and Bernoulli (1738), who argue that individuals evaluate outcomes in terms of utility rather than monetary value. Bernoulli (1738) illustrates this with a concave utility function, reflecting the principle of diminishing marginal utility of wealth. Based on this theory, von Neumann and Morgenstern (1944) formulated the Expected Utility Theory in which rational agents maximize expected utility under uncertainty. However, empirical evidence shows that investor behavior often deviates from these predictions.

These limitations are addressed by Kahneman and Tversky (1979) within their Prospect Theory, in which outcomes are evaluated relative to a reference point rather than in absolute terms. Regarding loss aversion, losses are perceived more strongly than equivalent gains, leading to risk aversion in gains and risk-seeking behavior in losses. This asymmetry helps explain notably

the disposition effect. Kahneman and Tversky (1992) later extended this framework through Cumulative Prospect Theory, refining the treatment of probabilities.

2.2.2. Mental accounting and framing effects

Starting with Prospect Theory, Thaler (1985) introduced mental accounting, explaining how individuals categorize, assess and track financial decisions in separate ‘cognitive accounts’. In fact, individuals tend to evaluate financial decisions separately rather than adopting a holistic view of their overall wealth or income. This can lead to inconsistent behavior even though both situations are similar. Thaler (1999) further shows that the attitude towards risk may vary depending on the domain of gains and losses. Closely related, framing effects imply that identical financial situations can lead to different decisions depending on how they are presented, potentially leading to biases such as the house money effect or loss aversion in investment choices (Tversky & Kahneman, 1981).

2.2.3. Cognitive and social drivers of investor behavior

Kahneman (2011) makes a distinction between two cognitive systems. The System 1 works automatically and uses heuristics, whereas the System 2 is slower and analytical. Investors are more subject to biases such as anchoring and overconfidence, because most financial decisions are driven by System 1. Although System 2 can mitigate these errors, it requires cognitive effort and is not always activated, which helps explain investor overreactions.

Beyond individual biases, social influences also shape investment decisions. Ajzen’s (1991) Theory of Planned Behavior shows that social norms and pressure can lead investors to conform to perceived best practices. Anderson and Holt (1997) experimentally demonstrate that informational cascades may emerge even when agents act rationally and that they may arise from low confidence or fear of nonconformity, thereby encouraging herding.

2.3. Investor herding behavior: Definitions and mechanisms

Investor herding occurs when investors align their decisions with those of others rather than relying only on private information. Some follow others because they believe other investors have better information (Bikhchandani & Sharma, 2001).

There is a difference between intentional and spurious herding. In intentional herding, investors deliberately imitate others due to reputational concerns (career risk), payoff externalities (liquidity or coordination benefits), or informational externalities (observing peers’ signals)

(Devenow & Welch, 1996). Spurious herding occurs when investors behave similarly because they react to the same information or preferences, without deliberate imitation. Spurious herding remains consistent with the rational investor hypothesis, while intentional herding may amplify mispricing and volatility (Merli & Roger, 2013). This distinction is important.

Where investors focus on particular assets, herding may enhance liquidity, but it may also contribute to speculative bubbles or panic-driven selling during crises. This understanding is crucial for both investors and regulators seeking to limit excessive volatility (Bikhchandani & Sharma, 2001).

Empirical research on herding has grown considerably over the decades. For example, Merli and Roger (2013) found significant mimetic behavior among French investors, showing that trading decisions often move collectively and not always in line with fundamentals. Similarly, Cauchies (2014) identifies herding patterns within limit order book dynamics.

2.4. Measurement of herding behavior

2.4.1. Asset-level herding: Lakonishok, Shleifer and Vishny measure (LSV)

The LSV metric is the most used measure of herding which quantifies, at the asset level for a given group of investors, whether buys and sells are concentrated on one side of the market more than it would be expected by ‘chance’ (Lakonishok et al., 1992).

Let $B_{j,t}$ and $S_{j,t}$, denote the number of net buyers and sellers of stock j in period t , with: $I_{j,t} = B_{j,t} + S_{j,t}$. The purchase intensity is: $p_{j,t} = \frac{B_{j,t}}{I_{j,t}}$

Let p_t denote the expected proportion of buyers in period t , typically computed as the average buy proportion across all stocks traded by the considered group during the same period. LSV herding measure for stock j in period t is then: $LSV_{j,t} = \left| \frac{B_{j,t}}{B_{j,t} + S_{j,t}} - p_t \right| - AF_{j,t}$. Or equivalently: $LSV_{j,t} = |p_{j,t} - p_t| - AF_{j,t}$

Under the null hypothesis of no herding, trading decisions are assumed to be independent, implying: $B_{j,t} \sim \text{Binomial}(I_{j,t}, p_t)$

Because the number of active traders $I_{j,t}$ varies across stocks and over time, deviations from p_t arise mechanically even without herding. The adjustment factor is:

$$AF_{j,t} = E \left[\left| \frac{B_{j,t}}{I_{j,t}} - p_t \right| \right] = E[|p_{j,t} - p_t|]. \text{ Or, } AF_{j,t} = \sum_{k=0}^{I_{j,t}} \binom{I_{j,t}}{k} (p_t)^k (1 - p_t)^{I_{j,t}-k} \left| \frac{k}{I_{j,t}} - p_t \right|$$

where k is the summation index (number of buyers). More details are provided in *Appendix A*.

LSV measures abnormal trade concentration by comparing the observed purchase intensity $p_{j,t}$ with the market benchmark p_t . The term $|p_{j,t} - p_t|$ captures the deviation from aggregate buying pressure, while subtracting $AF_{j,t}$ removes the expected binomial variation. Thus, $LSV_{j,t} = 0$ indicates independent trading, $LSV_{j,t} > 0$ indicates synchronous trading consistent with herding, and $LSV_{j,t} < 0$ indicates unusually balanced trading relative to random expectations.

Because p_t captures aggregate market-wide buying pressure, subtracting it prevents systematic liquidity shocks from being misinterpreted as herding. This is expressed in proportion.

In their original study, Lakonishok et al. (1992) apply the measure to quarterly holdings data of U.S. mutual funds over the period 1985-1989. Herding is therefore computed at the stock-quarter level. Investors are classified as being on the buy side (resp. sell side) of stock j over $[t - 1; t]$ if their number of shares held increases (resp. decreases) between $t - 1$ and t . Importantly, this classification is based on changes in the number of shares (not portfolio weights), and holdings are adjusted for corporate actions to avoid mechanical ‘buys’ or ‘sells’ driven by price movements or splits.

Lakonishok et al. (1992) report a mean herding of 0.027 across stock-quarters (median 0.001), implying that on average 2.7 percentage points (pp.) more managers trade on the same side of the market than would be expected under independent trading. Those results indicate statistically significant but economically moderate herding among US mutual funds.

This approach has been widely adopted and extended (e.g., to include trade sizes), but has some limitations: it may not distinguish intentional from spurious herding. Moreover, it cannot capture individual-level performance, and can be biased downward in small samples or when trading activity is quite low (Merli & Roger, 2013).

2.4.2. Investor-level herding: the Individual Herding Measure

While the LSV captures synchronized trading at the stock level, Merli and Roger (2013) introduce the Individual Herding Measure ($IHM_{i,t}$) to quantify herding behavior at the investor level. Unlike LSV, which measures whether a stock is traded disproportionately on one side of the market, the $IHM_{i,t}$ evaluates whether a given investor tends to trade in the same direction as the crowd on the stocks they actively trade during a given period.

The construction of the $IHM_{i,t}$ builds on LSV. A first step consists in distinguishing buy-side from sell-side herding through a signed version of LSV:

$$SLSV_{j,t} = \begin{cases} LSV_{j,t} & \text{if } p_{j,t} > p_t \\ -LSV_{j,t} & \text{if } p_{j,t} < p_t \end{cases}$$

If an investor buys a stock for which $SLSV_{j,t} > 0$, he trades in the same direction as the crowd and is considered to be herding on that stock. If he sells a stock with $SLSV_{j,t} < 0$, this also corresponds to herding, since the crowd is predominantly selling. Conversely, buying when $SLSV_{j,t} < 0$ or selling when $SLSV_{j,t} > 0$ corresponds to anti-herding (contrarian behavior). When $SLSV_{j,t} = 0$, trading is considered neutral. This framework ensures that the investor's behavior is evaluated relative to the prevailing aggregate trading direction.

The $IHM_{i,t}$ for investor i during period t is then defined as a value-weighted average of the signed stock-level signals:

$$IHM_{i,t} = \frac{\sum_{j=1}^J (\Delta n_{i,j,t} P_{j,t}) SLSV_{j,t}}{\sum_{j=1}^J |\Delta n_{i,j,t}| P_{j,t}}$$

where $\Delta n_{i,j,t} = n_{i,j,t} - n_{i,j,t-1}$ is the change in the number of shares held by investor i in stock j between $t - 1$ and t , and $P_{j,t}$ is the average price over the period $[t - 1; t]$. The denominator corresponds to the total euro value of transactions executed by the investor.

$IHM_{i,t}$ only considers stocks actively traded by the investor, ensuring that the measure reflects actual trading decisions rather than passive portfolio changes driven by price movements or unrelated assets. By construction, $IHM_{i,t} > 0$ indicates herding (trading in the same direction as the market), whereas $IHM_{i,t} < 0$ reflects contrarian behavior, and values close to zero indicate no systematic alignment with the crowd. Following Merli and Roger (2013), LSV used to

construct $LSV_{j,t}$ is set to zero when fewer than ten investors trade a stock, in order to limit small-sample distortions. The $IHM_{i,t}$ is computed at a monthly frequency.

Merli and Roger (2013) apply this approach to a large panel of 87,373 French investors over the period 1999-2006. They document significant and persistent individual herding, with substantial heterogeneity across investors and over time. Their results show that stronger herding is associated with lower performance and lower financial sophistication, while more experienced investors tend to herd less. By shifting the focus from stocks to investors, the $IHM_{i,t}$ provides a suitable framework to capture individual trading behavior and is therefore used as the main measure of herding in this thesis.

2.4.3. Alternative approaches and methodological debates

Herding can also be measured through the autocorrelation of order flows, which captures the persistence of trading decisions more directly than stock-level herding measures. Using order-book data at the daily stock level, Cauchies (2014) compares traditional trade- and return-based indicators with order-flow autocorrelation and analyzes the determinants of herding.

Debates around the measurement of herding exist. Despite its widespread use, LSV faces methodological criticism. Even though it proposes a ‘simple indicator’ to assess herding. Recent criticisms have shown that LSV suffers from systematic biases. More precisely, Jurkatis (2022) demonstrates through Monte Carlo simulations that LSV systematically underestimates true herding. The bias can range from 20 % to 100 % of the estimation target in some extreme scenarios, although typical values lie between 20 % and 80 % (which is still substantial), depending on the number of active investors and the proportion of herding traders. Moreover, he announces that this measure is in fact only a test of binomial overdispersion, and that other statistical methods offer better power and robustness properties. This must be considered to rigorously choose enhanced empirical tools to use in the following analysis.

2.5. Determinants of herding behavior

2.5.1. Individual characteristics

Merli and Roger (2013) found that poor past performance increases herding, which is consistent with the desire to avoid feeling regret. In contrast, investor sophistication, proxied in their study by derivatives use or the portfolio size, tends to attenuate herding. Given the specific context of the *Rallye Boursier*, sophistication is approximated using two complementary proxies

available in this dataset: success rates in some financial quizzes (organized during the competition to earn additional virtual capital), and portfolio diversification across international markets. Finally, anti-herding behavior, defined as trading against the crowd, generates higher ‘raw’ returns but lower risk-adjusted performance (e.g., Sharpe ratios) (Merli & Roger, 2013).

2.5.2. Trading environment factors

Trading environment characteristics condition the environment in which herding emerges. Using order-book data, Cauchies (2014) examines how stock-level characteristics affect the observed level of herding intensity. The relationship between volatility and herding depends on the context: while higher volatility may amplify informational uncertainty and incentivize herding, empirical evidence remains mixed. Blasco et al. (2009) document a positive association between herding and volatility, yet this relationship significantly varies with the measure and the approach employed.

Beyond volatility, inclusion in (national) indexes, greater trading activity, and higher cancellation rates (proportion of limit orders withdrawn before execution) are positively associated with measured herding. Stocks with greater visibility and stronger trading activity tend therefore to exhibit more herding. Moreover, the direction and magnitude of these relationships depend strongly on measurement choice (Cauchies, 2014).

2.5.3. Associated behavioral biases among ‘Rallye Boursier’ participants

Using the *Rallye Boursier* dataset, De Schrijver (2022) documents an inverse disposition effect, whereby participants tend to realize losses faster than gains. The intensity of this bias varies with diversification, trading activity, engagement, and age. While distinct from herding, it similarly highlights heterogeneity across investors, supporting the stratification by sophistication, age, and activity.

2.5.4. Momentum versus contrarian tendencies

Renard (2025)¹ investigates whether *Rallye Boursier* participants tend to follow market trends (momentum behavior) or deliberately trade against them (contrarian behavior), by comparing individual trades to benchmark ETF movements. To study heterogeneity, Renard (2025) segments participants by ‘past performance’ and ‘financial knowledge and engagement’ (similarly to De Schrijver (2022)). This segmentation reveals that both market conditions

¹ This master’s thesis was kindly shared directly by the author (Renard, 2025) before its official online publication.

(bullish / bearish periods) and investor profiles can strongly influence buy ratios. Less experienced or lower-performing participants tend to follow trends, whereas more sophisticated investors are more likely to adopt contrarian strategies (Renard, 2025).

Contrarian behavior may also take place in a group. When multiple investors simultaneously take positions against prevailing trends, this may result in ‘contrarian herding’ or anti-herding. As noted by Bikhchandani and Sharma (2001), such behavior can result from common information thus from shared reactions to the same market signals, rather than from explicit communication.

Apart from behavioral explanations, market frictions also play a role. Avramov, Chordia and Goyal (2006) show that short-term price corrections can be driven by liquidity effects, particularly in high-turnover or low-liquidity stocks. In this case, contrarian trading may reflect rational responses to temporary mispricing rather than purely behavioral dynamics. Distinguishing genuine information-based correction from ‘contrarian herding’ is therefore essential for interpreting correctly collective behavior.

2.5.5. Sociodemographic and cultural factors

Gender is one of the earliest dimensions studied in behavioral finance. In an article by Barber and Odean (2001), it is established that men trade more frequently and have greater overconfidence than women. The relationship with herding remains nuanced. While men may herd less due to greater confidence, their tendency to overreact to public information can lead to involuntary herding. In contrast, women, generally more risk- and loss-averse, may rely more on others’ actions when facing informational uncertainty. Korniotis and Kumar (2011) further highlight the role of financial knowledge.

Cultural factors also shape herding behavior. Zheng et al. (2021) and Lobão and Maio (2019) show that national culture affects both its intensity and direction. Building on Hofstede’s (2001) theory of cultural dimensions, they emphasize differences in individualism, uncertainty avoidance, and masculinity. Whereas collectivism and uncertainty avoidance cultures have investors seeking social validation and herd more, individualistic cultures promote autonomy and self-confidence, which leads to more ‘self-taken’ decisions, and therefore, reduce imitation (Zheng et al., 2021).

In Belgium, where French- and Dutch-speaking investors coexist, linguistic differences may reflect broader cultural distinctions that shape financial decision-making. Bellofatto (2017) shows that this linguistic proximity influences portfolio allocation patterns. French-speaking investors tend to overweight French stocks, while Dutch-speaking investors overweight Dutch stocks, reflecting language-based familiarity biases rather than differences in rationality. French and Walloon cultures share certain similarities for French-speaking investors, just as Dutch and Flemish environments do for Dutch-speaking investors. Cultural and linguistic proximity may thus structure informational environments and reference groups (Hofstede, 1980). This is consistent with Seitz et al. (2017), who find that such proximity increases herding in bank risk-taking decisions. Similarly, Andrikopoulos et al. (2017) document herding in Belgium and France but not in the Netherlands. Language stratification therefore provides a relevant framework to capture heterogeneity in herding.

Age constitutes another dimension, notably through digital exposure. Younger investors have deeply penetrated the digital world and are more impacted by social media than any other generation before. Hence, those young investors are highly exposed to digital information flows in their daily life. Algorithmic suggestions, online platforms and social media expose more and more young investors to an attention-driven herding where investment is made not only by their own or external information but completely replaced by trending searches and viral stocks (Zheng et al., 2021). The rise of online platforms such as Reddit and X (formerly Twitter) reinforces these dynamics (Nofsinger, 2022). Inversely, older investors rely more on their personal experience, leading to slower and less socially influenced decisions. Overall, younger investors tend to follow digital visibility, while older investors rely more on traditional information channels (Zheng et al., 2021).

Taken together, those show that herding is shaped by multiple interacting factors rather than a single bias. This complexity is addressed in the empirical analysis.

2.6. Digital catalysts of herding behavior

2.6.1. Attention-driven investing rise

Beyond demographic and cultural factors, declining attention also affects investor behavior. When information flows and algorithmic content filtering are abundant, investors increasingly focus on visible signals rather than on fundamentals. Barber and Odean (2008) show that individual investors tend to buy highly visible or popular assets rather than those with the strongest intrinsic value.

This dynamic is reinforced by the spread of financial narratives online. Shiller (2017) introduces the concept of narrative contagion, whereby viral financial stories circulate rapidly through social media. Combined with the fragmentation of attention described by Olicard (2021), these mechanisms increase cognitive reactivity and may strengthen herding tendencies among digitally connected investors. These attention constraints increase the diffusion of simplified and highly visible investment signals, which are increasingly transmitted through digital platforms.

2.6.2. Financial influencers ('finfluencers') role

While herding behavior has long existed in financial markets, its extent and speed have increased due to digitalization. Nowadays, social media platforms such as TikTok, YouTube, and LinkedIn disseminate investment opinions and strategies to large audiences (Hasanah et al., 2025).

These 'finfluencers' may influence trading decisions by sharing analyses, recommendations, or their own trading strategies. Platform algorithms can contribute even more to these signals by promoting viral content and concentrating investor attention on a limited set of assets (Wang & Wang, 2018). As noted by Wang and Wang (2018), 'these opinion leaders (i.e., gurus) may manipulate the information when it is disseminated to their followers'.

While digital factors amplify herding, individual and cultural determinants remain key drivers, as discussed in *Section 2.5.5*.

3. Methodology

This thesis relies on two complementary herding measures. The $IHM_{i,t}$ is computed at the investor level to study heterogeneity and its determinants, while the $LSV_{j,t}$ captures whether investors trade in the same direction at the stock level. This approach differs from order-flow autocorrelation methods (e.g., Cauchies, 2014), as it relies on executed transactions and focuses on directional trading behavior.

The analysis is based on the *Rallye Boursier*, an annual competition in which Belgian retail investors manage virtual portfolios under real market conditions (L'Echo, 2025).² Because participants operate in a common simulated environment while reacting to the same underlying market information, it is possible to assess whether investors trade independently or follow collective trading patterns.

3.1. Research gaps and contribution

Previous research shows that herding among individual investors can be economically meaningful but heterogeneous. Merli and Roger (2013) suggest that herding may persist over time and may vary with investor characteristics. Trading environments can also shape collective behavior (Cauchies, 2014). In the same *Rallye Boursier* context, other behavioral biases have already been documented, including the disposition effect (De Schrijver, 2022) and the coexistence of momentum and contrarian strategies (Renard, 2025).

Several gaps remain. Firstly, the measurement of herding can be improved, because $LSV_{j,t}$ is useful at the stock level, but it may underestimate herding due to sensitivity to sample size (Jurkatis, 2022), motivating the use of investor-level measures such as $IHM_{i,t}$. Then, the relationship between herding and performance remains ambiguous, since there is conflicting evidence on whether following or deviating from the crowd enhances individual outcomes (Merli & Roger, 2013). Furthermore, the recent literature emphasizes the role of attention, social interactions, and amplification mechanisms in financial markets (Barber & Odean, 2008; Shiller, 2017; Hasanah et al., 2025), which may influence trading behavior even if not directly observable in the dataset. Finally, demographic and cultural determinants of herding,

² A fee analysis on small trades (< €100, $n = 4,697$) indicates costs around €7.5 (+ 0.35% tax), consistent with prior editions (De Schrijver, 2022; Renard, 2025). Prior editions also imposed a 20-order daily limit, one buy and one sell order per stock per day, and a 10% cap per position.

particularly linguistic heterogeneity, remain relatively underexplored (Lobão & Maio, 2019; Bellofatto, 2017; Seitz et al., 2017).

This thesis addresses these gaps by combining stock- and investor-level herding measures and examining their relationship with investor characteristics and portfolio performance. It also assesses the robustness of the results across alternative specifications and sample definitions.

3.2. Research hypotheses

Building on the theoretical and empirical insights discussed above, this section formulates five hypotheses (H_A - H_E) on the determinants and consequences of herding among retail investors.

H_A: Existence of herding behavior among retail investors

H_{A0}: Retail investors trade independently.

H_{A1}: Retail investors exhibit statistically significant herding behavior.

This hypothesis reflects that trading decisions may be coordinated beyond what would be expected under independent behavior. Herding has been widely documented among both institutional and retail investors (Lakonishok et al., 1992; Bikhchandani & Sharma, 2001; Merli & Roger, 2013). Detecting herding provides a starting point to understand its determinants and consequences.

H_B: Relationship between financial sophistication and herding intensity

H_{B0}: Financial sophistication is not related to herding behavior.

H_{B1}: Higher financial sophistication is associated with lower herding intensity.

Merli and Roger (2013) show that the more experienced and informed investors are, the more they rely on personal judgment and less on imitation. Indicators such as portfolio diversification and quiz performance are therefore expected to be negatively related to herding intensity.

H_C: Age and herding behavior

H_{C0}: Age is not related to herding behavior.

H_{C1}: Younger investors exhibit stronger herding behavior.

Research by Korniotis and Kumar (2011) suggests that older and more experienced investors make more consistent and rational financial decisions, while younger participants tend to rely more on external signals and social influence. So, younger participants may be more exposed to social trading environments and attention cycles, leading to more herding.

H_D: Cultural and linguistic determinants of herding

H_{D0}: Cultural and linguistic differences are not related to herding behavior.

H_{D1}: Cultural and linguistic differences are associated with herding behavior.

Lobão and Maio (2019) argue that investors in more collectivist and uncertainty-averse cultures tend to herd more. Bellofatto (2017) identifies language-based familiarity biases among Belgian investors, highlighting how linguistic proximity impacts portfolio allocation decisions. Seitz et al. (2017) mention that cultural and linguistic proximity significantly increases herding in bank risk-taking decisions. These findings suggest that linguistic environment may shape imitation dynamics within investor groups, rather than reflecting systematic differences across groups.

H_E: Relationship between herding and portfolio performance

H_{E0}: Herding behavior is not related to raw portfolio performance.

H_{E1}: Higher herding is associated with lower portfolio performance.

Merli and Roger (2013) find that contrarian-herding investors achieve higher gross returns, although this advantage disappears after adjusting for risk. This thesis therefore examines whether stronger individual herding intensity is associated with lower performance.

Together, these hypotheses aim to clarify how behavioral, demographic, and cultural factors influence individual investor herding and how such behavior relates to portfolio performance.

3.3. Data construction and cleaning procedures

The empirical analysis relies on four confidential datasets from the *Rallye Boursier*, covering the period from November 2019 to January 2020. All monetary values are expressed in euros.

The dataset ‘Players’ contains investor characteristics, engagement indicators (*LOGINS_i*), quiz results, and final portfolio performance. ‘All Orders’ records all submitted orders (*USER_ID*, *ISIN*, date, quantity, direction, and execution status). ‘Daily Prices Stocks’ provides daily stock prices and trading volumes, while ‘Stocks’ includes firm-level information such as country, industry classification, index membership, and descriptive characteristics of listed firms.

The raw data were available in SAS format and imported into R for processing. Constructing the final dataset required merging these sources across three dimensions: investor, stock, and time. The data were then modified into a consistent investor-stock-month structure.

Only executed orders are retained. Transaction-level data are aggregated to compute net trading positions (buy minus sell) at the investor-stock-month level. Daily stock prices are aggregated into monthly averages to match the frequency of the herding measures.

The resulting dataset combines trading activity, price information, and investor characteristics, and serves as the basis for all empirical analyses.

3.3.1. Implementation overview (data pipeline)

To guide the reader, the construction of the dataset follows a structured pipeline. Firstly, executed orders are aggregated at the investor-stock-month level to compute net positions ($\Delta n_{i,j,t} = \text{buys} - \text{sells}$). Secondly, stock-level herding measures ($LSV_{j,t}$ and $SLSV_{j,t}$) are computed using these aggregated flows. Thirdly, investor-level herding ($IHM_{i,t}$) is constructed by combining these stock-level signals with investor trading positions, using either value weights ($IHM_{i,t}^{pw}$) or equal weights ($IHM_{i,t}^{ew}$). All computations are implemented in R using grouped aggregations at the investor-stock-month level.

3.3.2. Sample selection and definition of active investors

The analysis focuses on active investors. Accounts with no platform activity ($LOGINS_i = 0$) or no stock orders ($ORDERS_STOCKS = 0$) are excluded, as herding requires observable trading.

The dataset is structured at the investor-month level for November 2019, December 2019, and January 2020. An observation (investor i in month t) is retained only if at least 10 BUY orders were executed during that month ($LAST_STATUS = 'X'$).

This threshold ensures that the herding measure is computed on sufficiently active trading behavior and reduces noise from sporadic participation.³ As a robustness check, the same pipeline is also implemented with a lower threshold ($MIN_BUYS = 5$).

The final panel consists of investor-month observations, with each investor contributing up to three observations (three months in the dataset) depending on activity.

3.3.3. Data cleaning and treatment of specific cases

Only executed orders are retained, excluding cancelled or rejected orders.

³ This threshold introduces a moderate selection toward more active investors.

Trading activity is measured using changes in the number of shares held rather than valuation changes. This choice avoids distortions due to price fluctuations or stock splits. These variations are valued using average monthly prices to reflect trading decisions rather than market movements.

Observations with missing key identifiers (USER_ID or ISIN) are removed. Demographic variables are screened to eliminate implausible values. Ages below 18 or above 90 are treated as outliers and excluded, leading to the removal of 4 observations. These cases are negligible and do not affect the results. Only when necessary, winsorization is done.⁴

The initial stock universe contains 342 ISINs. To ensure a consistent trading universe, the sample is restricted to equities using the stock reference file (TYPE = 'STOCK', STOCK_TYPE ∈ {share, trust}) (*Appendix B.*). This restriction reduces the universe to 338 securities, excluding four inactive or delisted stocks.⁵

3.3.4. Construction of the herding measure

Herding is measured using two complementary approaches: a stock-level measure ($LSV_{j,t}$) and an investor-level measure ($IHM_{i,t}$).

At the stock level, the LSV measure is computed for each stock j and month t by comparing the proportion of buyers $p_{j,t}$ to the market-wide proportion p_t . The measure is constructed following Lakonishok et al. (1992) and is only computed for stock-month observations with a sufficient number of active investors. A signed version, $SLSV_{j,t}$, is used to distinguish between excess buying and excess selling.

At the investor level, the Individual Herding Measure aggregates these stock-level signals across the trades of each investor. The value-weighted specification is defined as:

$$IHM_{i,t}^{vw} = \frac{\sum_j \Delta n_{i,j,t} P_{j,t} SLSV_{j,t}}{\sum_j |\Delta n_{i,j,t}| P_{j,t}}$$

where $\Delta n_{i,j,t}$ denotes the net change in shares (buys minus sells) and $P_{j,t}$ is the average monthly price of stock j . The numerator captures alignment with aggregate trading, while the denominator scales by total trading activity.

⁴ Winsorization caps extreme values at the 1st and 99th percentiles (*see Section 4.4 and Table II.*)

⁵ This restriction is applied before constructing investor-level variables to ensure a consistent and economically meaningful trading universe. Dropped stocks: BE0974352842, FR0013399474, BE0003700144, FR0011790542.

By construction, the stock-level herding measure satisfies $LSV_{j,t} \in [-AF_{j,t}, 1]$,⁶ while its signed version satisfies $SLSV_{j,t} \in [-1, 1]$. At the investor level, the Individual Herding Measure satisfies $IHM_{i,t} \in [-1, 1]$. In practice, however, observed values are typically much smaller in magnitude, indicating moderate deviations from independent trading.

To facilitate the interpretation, a numerical illustration of the construction of $LSV_{j,t}$, $SLSV_{j,t}$ and $IHM_{i,t}$ is provided in *Appendix C*. This example uses three assets and three periods to guide the reader through each step of the computation.

3.3.4.1. Value-weighted ($IHM_{i,t}^{vw}$) and equal-weighted ($IHM_{i,t}^{ew}$) specifications

Two versions of the $IHM_{i,t}$ are considered. The baseline specification is value-weighted ($IHM_{i,t}^{vw}$), where each trade is weighted by its monetary value. This reflects the economic importance of trades, as larger positions have a stronger impact on portfolio outcomes. As a robustness check, an equal-weighted ($IHM_{i,t}^{ew}$) version is also constructed. In this case, each trade receives the same weight, regardless of its size: $IHM_{i,t}^{ew} = \frac{1}{N_{i,t}} \sum_j \text{sign}(\Delta n_{i,j,t}) * SLSV_{j,t}$ where $N_{i,t}$ is the number of traded stocks.

The $IHM_{i,t}^{ew}$ measure captures directional alignment only, while the $IHM_{i,t}^{vw}$ measure incorporates both direction and economic magnitude.

In practice, both measures are computed in R by aggregating transaction-level data at the investor-stock-month level. The $IHM_{i,t}^{vw}$ version uses price-weighted flows, while the $IHM_{i,t}^{ew}$ version relies on simple averages of trade signs.

The $IHM_{i,t}^{vw}$ specification is retained as the baseline because it better reflects capital allocation decisions, while the $IHM_{i,t}^{ew}$ specification is used to verify that results are not driven by a small number of large trades. In the value-weighted measure, larger trades have greater influence.

3.3.5. Construction of explanatory variables

Age_i is computed at the beginning of the competition (November 2019) using date of birth.

⁶ Although LSV measure is often interpreted as a non-negative indicator of herding intensity, it may take slightly negative values in practice. This occurs when observed imbalance in trading is smaller than expected deviation under the null of independent trading, captured by the adjustment factor. Negative values therefore reflect unusually balanced trading rather than the absence of herding.

Language is proxied using the registration platform. Investors registering through *L'Echo* are classified as French-speaking, while those registering through *De Tijd* are classified as Dutch-speaking. A binary variable $French_i$ equals 1 for French-speaking investors and 0 otherwise.

Financial $Sophistication_i$ is measured using a composite index combining two dimensions.

The first component is a quiz score:

$$QuizScore_i = \frac{QUIZ_OK_i}{QUIZ_OK_i + QUIZ_NOK_i}$$

This variable captures financial knowledge. For example, an investor with 8 correct answers and 2 incorrect ones obtains a $QuizScore_i$ of $\frac{8}{(8+2)} = 0.8$. Values close to 1 indicate a high level of financial knowledge, while values close to 0 indicate a low level.

The second component is an international diversification index:

$$IntlDiv_i = \frac{\text{Number of countries traded by } i}{\text{MAX number of countries traded by a single investor in the sample}}$$

This variable captures the geographical breadth of trading activity. For example, if an investor trades in 4 countries and the most internationally diversified investor in the sample trades in 12 countries, then $IntlDiv_i = \frac{4}{12} = 0.33$. Values close to 0 indicate limited international diversification, while values close to 1 correspond to investors whose trading activity is spread across many countries. In the sample, investors trade in 5.55 countries on average, with a median of 5. Most investors trade in 4 to 7 countries, while only 8 investors trade in a single country. The maximum observed value is 12 countries, which is used for normalization.⁷

The final $Sophistication_i$ index is computed as the average of both components:

$$Sophistication_i = \frac{QuizScore_i + IntlDiv_i}{2}$$

This combines knowledge (quiz performance) and behavior (diversification). For example, if $QuizScore_i = 0.8$ and $IntlDiv_i = 0.33$, then $Sophistication_i$ equals $\frac{0.8 + 0.33}{2} = 0.565$. Intuitively, an investor is considered more sophisticated if they both understand financial concepts and adopt diversification strategies. Both components are normalized between 0 and 1 and receive equal weight, ensuring a balanced and interpretable measure.⁸

⁷ While 13 countries are available in the cleaned stock universe, the maximum number traded by a single investor is 12. Normalization therefore relies on the sample maximum.

⁸ Alternative weighting schemes could be considered. The 50-50 specification is retained for simplicity.

Trading activity is measured as the number of executed stock orders in month t . A logarithmic transformation $\log(\text{Trades}_{i,t})$ is used in the regressions to capture marginal impacts.

Investor engagement is proxied by the number of platform logins (LOGINS_i), defined as the number of times investor i accessed the trading platform over the competition period.⁹

Portfolio performance is measured as the cumulative return over the full competition period (November 2019 to January 2020) and is therefore observed once per investor. In contrast, herding intensity is computed at the investor-month level. To make the two variables comparable, the monthly herding measure ($\text{IHM}_{i,t}^{vw}$) is averaged across months for each investor. This yields an investor-level measure of average herding, denoted $\overline{\text{IHM}}_i^{vw}$, which is then used to analyze the relationship between herding and performance.¹⁰

Although the original dataset covers three months, November corresponds to a portfolio formation phase during which most investors are net buyers. The main analyses therefore focus on December and January, where trading behavior is more balanced.¹¹

3.4. Empirical strategy

This section explains how the hypotheses are tested using the constructed dataset. The analysis combines investor-level panel regressions and stock-level measures, with a focus on explaining how herding is measured, modeled, and interpreted.

3.4.1. Panel data structure and sample definition

The empirical analysis relies on a short panel dataset where the unit of observation is investor i in month t . Although the dataset initially covers three months (November 2019 to January 2020), November corresponds to a portfolio formation phase during which most investors are net buyers. As a result, herding measures are mechanically close to zero at the stock level. For this reason, the core analysis focuses on December 2019 and January 2020, where both buying and selling decisions are observed and herding becomes informative.

The final sample contains 2,693 investor-month observations, corresponding to 2,208 unique investors. Each investor contributes one or two observations depending on their activity.

⁹ A binary login indicator ($\text{logged}_{i,t}$), equal to 1 if investor i logs in at least once in month t (0 otherwise), was initially constructed but excluded, as the activity threshold ($\text{MIN_BUYS} = 10$) already ensures active participation.

¹⁰ Given that the analysis covers two months, using the median instead of the mean would yield identical results.

¹¹ This feature is examined in detail in *Section 4.4.6.*, with further technical evidence reported in *Appendix E.*

Herding is measured at two levels. At the stock level, $LSV_{j,t}$ and $SLSV_{j,t}$ capture synchronization in trading across investors. Stock-level signals ($SLSV_{j,t}$) are then combined with each investor's own trades to construct investor-level herding measure ($IHM_{i,t}^{vw}$). For each investor, the measure aggregates the $SLSV_{j,t}$ of the stocks they trade, weighted by their transactions size.

Stock-level herding measures are computed only when the number of active investors $I_{j,t}$ is sufficiently large. In December-January, 575 out of 671 stock-month observations are measurable under the $I_{j,t} \geq 10$ condition ($\frac{575}{671} = 85.7\%$). The full dataset (November-January) initially includes 1,198 observations (*Tables I2. and I3.*).

This structure ensures that the analysis focuses on informative trading while limiting small-sample distortions.

3.4.2. Model specification and key methodological choices

The baseline empirical specification aims to explain individual herding intensity as a function of investor characteristics and trading behavior. The model is estimated using the following panel regression:
$$IHM_{i,t}^{vw} = \alpha + \beta_1 Sophistication_i + \beta_2 Age_i + \beta_3 French_i + \beta_4 \log(Trades_{i,t}) + \lambda_t + \varepsilon_{i,t}$$

where $IHM_{i,t}^{vw}$ is the value-weighted herding measure, λ_t represents month fixed effects, and $\varepsilon_{i,t}$ is the error term.

Three key methodological choices underlie this specification: the definition of active investors, the treatment of trading activity, and the use of fixed effects (FE).

3.4.2.1. Eligibility rule: MIN_BUY threshold

The panel includes only investor-month observations where at least 10 BUY orders are executed. This threshold ensures that herding is measured on sufficiently active trading.

Buy orders are used rather than total trades because they more clearly reflect the initiation of new positions. Sell orders can arise from various reasons (e.g. profit-taking or position adjustments), making them less directly comparable across investors.

An alternative definition based on total trades ($MIN_TRADES = 10$) is also considered. In practice, both definitions lead to identical samples in this dataset. The $MIN_BUY = 10$ rule is therefore retained as the baseline.

3.4.2.2. Trading activity specification: log transformation

Trading activity is measured as the number of trades executed by investor i in month t . Two specifications are considered: the level of trades and its logarithmic transformation.

The logarithmic specification is retained in the baseline model because it captures diminishing marginal impacts. For instance, an increase from 2 to 10 trades is likely to affect behavior more strongly than an increase from 52 to 60 trades. Both specifications are reported in robustness checks (*Tables 14. and 15.*).

3.4.2.3. Fixed effects and clustering

Month fixed effects (λ_t) are included to control for common shocks affecting all investors, such as market conditions or competition dynamics. As a result, identification relies on cross-sectional differences between investors within each month.

Investor fixed effects are not included because they would absorb all time-invariant variables (Age_i , $Sophistication_i$, $French_i$). Given the short time dimension of the panel, such a specification would also provide limited variation for estimation.

Standard errors (SE) are clustered at the investor level (USER_ID) to account for correlation across observations belonging to the same investor. Clustering at the month level is not appropriate due to the limited number of time periods.

Overall, this empirical strategy allows identifying how individual characteristics and trading are associated with herding intensity, while controlling for common market conditions.

3.4.3. Hypothesis testing framework (Tests for each hypothesis)

This subsection explains how each hypothesis is empirically tested. The analysis combines descriptive statistics, panel regressions, and cross-sectional investor-level regressions. The main dependent variable is the value-weighted herding measure $IHM_{i,t}^{vw}$. The equal-weighted version $IHM_{i,t}^{ew}$ is used as a robustness check.

3.4.3.1. HA: Existence of herding ($IHM \neq 0$)

Hypothesis HA tests whether retail investors exhibit herding behavior. Herding intensity is measured using $IHM_{i,t}^{vw}$. Under the null hypothesis, $H_0: E[IHM_{i,t}^{vw}] = 0$

This hypothesis is tested using a one-sample two-tailed t-test of $\overline{IHM}^{vw}$ against zero.

In addition, descriptive statistics are used to assess the distribution of herding across investors. In particular, the shares of investor-month observations with $IHM_{i,t}^{vw} > 0$ (herding), $IHM_{i,t}^{vw} < 0$ (contrarian behavior), and $IHM_{i,t}^{vw} = 0$ are examined.

These elements allow evaluating both the existence and the prevalence of herding.

3.4.3.2. HB, HC, HD: Determinants of herding (panel regressions)

Hypotheses HB, HC, and HD examine whether herding is related to investor characteristics. These hypotheses are tested using the panel regression introduced in *Section 3.4.2*:

$IHM_{i,t}^{vw} = \beta_0 + \beta_1 \text{Sophistication}_i + \beta_2 \text{Age}_i + \beta_3 \text{French}_i + \beta_4 \log(\text{Trades}_{i,t}) + \lambda_t + \varepsilon_{i,t}$ where $IHM_{i,t}^{vw}$ is the dependent variable, λ_t denotes month FE, and SE are clustered at the investor level.¹²

Each hypothesis is evaluated through the corresponding coefficient:

HB: β_1 (Sophistication); HC: β_2 (Age); HD: β_3 (Language).

The sign, magnitude, and statistical significance of these coefficients determine whether the hypotheses are supported.

Additional checks are performed to better understand the role of language (HD). These include comparisons of mean herding across language groups, two-sample t-tests, and alternative specifications including only the language dummy or interaction terms. These complementary tests allow distinguishing unconditional differences from conditional relationships.

$IHM_{i,t}^{vw}$ and alternative specifications are used as robustness checks to verify that results are not driven by the weighting scheme or model specification.

¹² Variable $\text{Logged}_{i,t}$ was initially included but dropped due to collinearity and is therefore not retained in the baseline model.

3.4.3.3. HE: Herding and performance (using cross-sectional investor-level regressions)

Hypothesis HE examines whether herding is associated with portfolio performance.

The analysis is conducted using an investor-level cross-section, because performance is observed once per investor. The monthly herding measure is first averaged across months:

$$\overline{IHM}_i^{vw} = \frac{1}{T_i} \sum_t IHM_{i,t}^{vw} \left(= \frac{1}{2} \sum_t IHM_{i,t}^{vw} \right) \text{ Since there are 2 months in the sample, } T_i = 2$$

The final specification is: $PERFORMANCE_i = \alpha_0 + \alpha_1 \overline{IHM}_i^{vw} + \alpha_2 Age_i + \alpha_3 Sophistication_i + \alpha_4 French_i + \alpha_5 LOGINS_i + \varepsilon_i$

Trading activity is not included in the cross-sectional specification, as it is defined at the investor-month level and cannot be directly used in a cross-section without aggregation. Additionally, such aggregation may introduce endogeneity concerns.

The coefficient α_1 captures the relationship between herding and performance. Under the null hypothesis: $H_0: \alpha_1 = 0$

Inference is based on White heteroscedasticity-robust standard errors. A winsorized version of performance is also used as a robustness check to account for extreme values.

The sign, magnitude, and stability of α_1 provide the empirical test of the hypothesis. A negative and significant coefficient indicates that stronger herding is associated with lower raw portfolio performance.

3.4.4. Inference and model evaluation

Model evaluation combines coefficient significance, economic amplitude, and some (fit) statistics that indicate how well the model explains the data. Because the dependent variable reflects noisy individual trading decisions, fit statistics should be interpreted with caution.

Reported statistics include R^2 , Adjusted R^2 , and, for panel models with fixed effects, the Within R^2 . Let $X_{i,t}$ denote the vector of explanatory variables ($Sophistication_i$, Age_i , $French_i$, and $\log(Trades_{i,t})$). If the model is written as $IHM_{i,t}^{vw} = X_{i,t} \beta + \lambda_t + \varepsilon_{i,t}$, the Within R^2 measures the fraction of variation in $IHM_{i,t}^{vw} - \overline{IHM}_t^{vw}$ explained by the transformed regressors $X_{i,t} - \bar{X}_t$, where bars denote month averages. In practice, it captures how much of the cross-sectional variation in herding across investors, within a given month, is explained by the covariates.

In short panels, the Within R^2 is the most informative measure, as the overall R^2 can be mechanically influenced by fixed effects. Interpretation focuses on the sign, magnitude, and robustness of coefficients rather than predictive accuracy. Additional methodological details are provided in *Appendix D*.

3.5. Robustness strategy & checks

To assess the stability of the main findings, several robustness checks are conducted.

Firstly, alternative sample definitions are considered by comparing the baseline specification ($\text{MIN_BUYS} = 10$) with a less restrictive threshold ($\text{MIN_BUYS} = 5$), and with an alternative eligibility rule based on total executed trades ($\text{MIN_TRADES} = 10$).

Secondly, alternative specifications are tested. These include winsorizing portfolio performance at the 1st and 99th percentiles, replacing the value-weighted herding measure with its equal-weighted version, and comparing level and logarithmic specifications of trading activity.

Thirdly, a sensitivity check concerns the stock-level construction of herding. Because November 2019 corresponds to the portfolio formation phase, $LSV_{j,t}$ is mechanically close to zero in that month. The main analysis therefore focuses on December and January, with additional methodological remarks on this point reported in *Appendix E*.

The detailed results of all robustness checks are presented in *Section 4.4*.

4. Results

4.1. Descriptive evidence on herding

Descriptive statistics are reported in *Table 1*. The sample is composed of 2,693 investor-month observations corresponding to 2,208 unique investors.¹³

The average level of herding is positive, indicating a moderate but systematic tendency for investors to follow aggregate trading flows (*see Table 1*). $IHM_{i,t}^{ew}$ remains positive on average in both months and increases slightly in January (*Figure H1*). Trading activity is right-skewed, as the mean number of trades exceeds the median. Trading is moderately diversified but remains concentrated on a limited number of markets, particularly within Europe.¹⁴ Further details on sample construction and activity thresholds are provided in *Appendix F*.

Table 1. Descriptive statistics

<i>Definition</i>	<i>Variable</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
3.3.4.	$IHM_{i,t}^{vw}$	0.0171	0.0160	0.0300	-0.1969	0.1560
3.3.4.	$IHM_{i,t}^{ew}$	0.0157	0.0153	0.0279	-0.1978	0.1561
3.3.5.	Age_i	45.47	46	18.58	18	90
3.3.5.	$Sophistication_i$	0.5569	0.5670	0.1270	0.0769	0.9075
3.3.5.	$French_i$	0.2796	0	0.4489	0	1
3.3.5.	$Trades_{i,t}$	28.35	23	20.33	10	295
3.3.5. & 3.4.2.2.	$Log(Trades_{i,t})$	3.1838	3.1355	0.5340	2.3026	5.6870

¹³ The panel is short and unbalanced: most investors are observed in a single month (1,723), while a smaller fraction appears in both months (485).

¹⁴ The full set of countries includes Belgium, France, the Netherlands, the United States, Canada, Switzerland, the United Kingdom, Iceland, Ireland, Luxembourg, Germany, China, and South Africa.

4.1.1. Distribution of IHM (HA)

$\overline{IHM}^{vw}$ is positive (0.0171) in the baseline panel ($N = 2,693$), providing initial evidence of herding behavior. A one-sample t-test strongly rejects the null hypothesis of independent trading ($t = 29.624$, $p < 0.001$). The 95% confidence interval $[0.01597; 0.01824]$ excludes zero, confirming the presence of positive average herding.

As shown in *Figures H2. and H3.*, $IHM_{i,t}^{vw}$ is centered above zero, with a majority (72.15%) of investor-month observations exhibiting positive herding. This pattern strongly supports HA_1 , indicating that investors tend to align their trades with aggregate market activity.

4.1.2. Stock-level herding (LSV Overview)

Herding is first examined at the stock-month level using the $LSV_{j,t}$ measure. Over the informative trading period (December-January), 575 out of 671 stock-month observations are measurable (85.7%), while 14.3% are excluded due to insufficient investor participation (*Table I2.*). This indicates that the $LSV_{j,t}$ measure can be computed for the large majority of stock months, ensuring that the analysis captures a representative share of overall trading activity (*Figures H4. and H5.*).

November is excluded, as it corresponds to a portfolio formation phase during which most investors are net buyers, mechanically leading to $LSV_{j,t}$ values close to zero. The analysis therefore focuses on December and January.

Average $LSV_{j,t}$ values are 0.0299 in December and 0.0405 in January, indicating a non-negligible level of coordinated trading across stocks. $SLSV_{j,t}$ shows a roughly balanced distribution between buy and sell synchronization (*Table I3.*), suggesting that herding is not systematically directional.

These magnitudes are consistent with prior studies. Lakonishok et al. (1992) report $\overline{LSV}$ values of approximately 2.7% for institutional investors, while the estimates observed here range between 3% and 4%. Although not directly comparable due to differences in data and sample composition, these results suggest that the level of herding observed in this study is of a similar order of magnitude.

These descriptive results naturally lead to the regression analysis.

4.2. Determinants of herding (HB, HC, HD)

4.2.1. Baseline panel regressions (MIN_BUYS = 10)

Table 2. reports the results of the fixed-effects panel regression of $IHM_{i,t}^{pw}$ based on 2,693 investor-month observations, with standard errors clustered at the investor level (USER_ID).

Financial sophistication is positively associated with herding intensity, indicating that more sophisticated investors tend to align more closely with aggregate trading. This finding contrasts with HB₁ and suggests that herding may reflect common information processing rather than imitation driven by lack of knowledge.

Similarly, the positive relationship with age contradicts HC₁ and indicates that older investors follow aggregate trading patterns slightly more closely, possibly relying more on established heuristics or widely observed signals.

No significant difference is observed between French- and Dutch-speaking investors. As a result, HD₁ is not supported, suggesting that language does not play a major role in explaining herding behavior, ceteris paribus (Figure H6.).

Trading activity is strongly related to herding, with more active investors exhibiting greater synchronization with market flows (Figure H7.).

Table 2. Baseline determinants of herding ($IHM_{i,t}^{pw}$), MIN_BUYS = 10

Variable	Coefficient	Std. Error
<i>Sophistication_i</i>	0.0105**	(0.0050)
<i>Age_i</i>	9.24 * 10 ⁻⁵ ***	(0.00003)
<i>French_i</i>	-0.0021	(0.0013)
<i>log(Trades_{i,t})</i>	0.0064 ***	(0.0012)
Month fixed effects	Yes	-
Observations	2,693	-
R ²	0.02534	-
Adjusted R ²	0.02353	-
Within R ²	0.02290	-

Notes: Standard errors clustered at the investor level (USER_ID).

*, **, *** denote significance at the 10%, 5%, and 1% levels.

4.2.2. Economic implications

Trading activity has the strongest economic impact on herding. An increase in trading activity is associated with a rise in $IHM_{i,t}^{vw}$, which imply that more active investors are more likely to follow aggregate market behavior. For example, a rise in trading activity from 10 to 30 trades corresponds to a change of approximately one unit in $\log(Trades_{i,t})$, since $\log_e(30) - \log_e(10) = 1.0986$. Given the estimated coefficient, this is associated with an increase of about 0.007 in $IHM_{i,t}^{vw}$ ($0.0064 * 1.0986 \approx 0.007$), corresponding to roughly 37% of the sample mean.¹⁵

Although the coefficient on financial sophistication is larger in absolute terms, its economic impact is limited due to the low variation of the variable. Financial sophistication varies within a relatively narrow range, so even meaningful changes translate into small variations in herding. A typical variation of 0.1 leads to a much smaller change in herding ($0.0105 * 0.1 \approx 0.001$). Overall, despite a larger coefficient, the economic effect of financial sophistication remains limited relative to trading activity.

Age is also positively associated with herding, but its economic effect remains modest compared to trading activity. For Age_i , a 25-year difference between two investors increases $IHM_{i,t}^{vw}$ by 0.002309 (ceteris paribus), obtained by scaling the estimated coefficient. Given a mean value of $\overline{IHM}^{vw}$ of 0.0171, this represents approximately 13.5% of the average level, and about 7.7% of a one-SD change ($SD = 0.0300$).¹⁶

Additional exploratory analyses suggest that herding may be stronger when investors trade highly popular stocks, potentially reflecting attention concentration effects within the simulation environment (*Figures H8. and H9.*).¹⁷

More broadly, herding is not only statistically significant but also economically meaningful, particularly for highly active investors.

4.2.3. Model fit and explanatory power

In the baseline specification, R^2 is about 0.025 and Within R^2 is about 0.023, which is consistent with noisy individual-level data. These values are lower than those reported by Merli and Roger

¹⁵ Obtained from the estimated coefficient on $\log(Trades_{i,t})$ ($\beta \approx 0.0064$). Relative to the mean herding level (≈ 0.0171), this corresponds to $\approx 37\%$ ($0.0064 / 0.017103 \approx 0.374$).

¹⁶ Obtained by multiplying the estimated Age coefficient ($\beta_{Age} = 9.235095 \times 10^{-5}$) by a 25-year difference: $9.235095 \times 10^{-5} * 25 \approx 0.002309$. Relative to the mean herding level (≈ 0.0171), this corresponds to $\approx 13.5\%$ ($0.002309 / 0.017103 \approx 0.135$), and $\approx 7.7\%$ of one-SD ($0.002309 / 0.0300 \approx 0.077$).

¹⁷ With the top 20% most traded stocks, the relationship remains positive and significant (*Appendix G.*).

(2013) ($R^2 \approx 0.25$, from their ‘Table 7’), but not directly comparable due to aggregation differences (fund-level vs. investor-level regressions) and the higher idiosyncratic variation in trading decisions. Conversely, Barber and Odean (2001) report similarly modest explanatory power in retail investor settings ($R^2 \approx 1.26\%$, from their ‘Table III’), suggesting that relatively low fit is common in this empirical context. Additional details are provided in *Appendix D*.

4.3. Herding and portfolio performance (HE)

This section focuses on raw portfolio performance. A risk-adjusted measure was explored but discarded, as its reliable construction would have required a full day-by-day reconstruction of each investor’s portfolio, including cash holdings and transaction timing, potentially introducing measurement error.¹⁸

The relationship between herding intensity and portfolio performance is analyzed using an investor-level regression ($N = 2,143$). *Table 3.* shows that herding is negatively and significantly associated with portfolio performance across all specifications. In the baseline model, the coefficient on $\overline{IHM}_i^{vw}$ is -0.1476 and decreases to -0.2166 when controlling for investor engagement ($LOGINS_i$), which is retained as the preferred specification. Therefore, when this effect is taken into account, the negative relationship between following the market and performance becomes more pronounced.

Economically, a 0.1 increase in $\overline{IHM}_i^{vw}$ corresponds to a decrease of about 1.5 pp. in performance ($\approx 0.1 * -0.1476$), with a larger effect in the preferred model. Moreover, a one-SD increase in $\overline{IHM}_i^{vw}$ (0.0281) is associated with a decrease in performance of 0.00609. Moving from zero herding to the sample mean of $\overline{IHM}_i^{vw}$ (0.0169) corresponds to a decrease of 0.00365 (*Figure H10.*). These results are economically meaningful given the short investment horizon (Dec-Jan). The model explains approximately 15% of the variation in investor performance.¹⁹

Overall, herding investors tend to achieve lower returns, likely because they enter positions after information has already been incorporated into prices. This negative relationship remains robust across specifications, including when raw performance is winsorized, and supports HE.

¹⁸ Portfolio performance is measured over Nov-Jan, while herding is averaged over Dec-Jan. Since November corresponds to a formation phase with limited herding variation, this likely attenuates the estimated relationship.

¹⁹ Obtained by applying the estimated coefficient on average herding ($\beta_{\overline{IHM}_i^{vw}} \approx -0.2166$) to observed herding variations. A one-SD increase (0.0281) implies $-0.2166 * 0.0281 \approx -0.0061$, corresponding to a decrease of 0.61 pp. in performance. Moving from zero herding to sample mean (0.0169) implies $-0.2166 * 0.0169 \approx -0.00365$.

Several behavioral mechanisms may explain this pattern. Informational cascade theories suggest that investors may infer information from aggregate market movements or widely shared signals when private information is limited, potentially leading to synchronized trading after prices have already adjusted (Anderson & Holt, 1997). Herding may therefore reflect attention-driven behavior or common reactions to public information rather than direct trade replication (Bikhchandani & Sharma, 2001; Devenow & Welch, 1996). Although the simulation may have allowed delayed access to some portfolio information, exact real-time copying was not possible, suggesting that the observed synchronization primarily reflects shared information and market signals.

The conclusions are consistent with Merli and Roger (2013), who show that investors trading against the crowd achieve higher returns. Attention-driven mechanisms may further reinforce these dynamics: retail investors tend to concentrate on highly visible stocks (Barber & Odean, 2008), while narratives can amplify collective behavior (Shiller, 2017). The GameStop episode illustrates how retail attention can generate price pressure and subsequent losses (Grant, 2021; Pomparelli, 2024), although the underlying mechanisms extend beyond this specific case.

Table 3. Herding and portfolio performance

Dependent variable	Baseline model (without engagement control)		Final model (with engagement control)	
	Performance	Performance (winsorized)	Performance	Performance (winsorized)
$\overline{IHM}_i^{vw}$	-0.1476*** (0.0437)		-0.2166*** (0.0422)	
$\overline{IHM}_i^{vw}$ (with winsorized performance)		-0.1395*** (0.0420)		-0.2111*** (0.0405)
Constant	0.0010 (0.0058)	0.0049 (0.0051)	-0.0043 (0.0058)	-0.0002 (0.0050)
Age _i	0.0002*** (0.000059)	0.0002*** (0.000056)	0.000076 (0.000058)	0.000073 (0.000055)
Sophistication _i	0.1231*** (0.0099)	0.1157*** (0.0088)	0.0776*** (0.0096)	0.0721*** (0.0088)
French _i	-0.0102*** (0.0026)	-0.0098*** (0.0025)	-0.0115*** (0.0025)	-0.0110*** (0.0024)
LOGINS _i			0.0008*** (0.000067)	0.0008*** (0.000061)
Observations	2,143	2,143	2,143	2,143
R ²	0.08980	0.09119	0.15021	0.15527
Adjusted R ²	0.08809	0.08949	0.14822	0.15330

Notes: White heteroscedasticity-robust SE in parentheses. Performance is winsorized at the 1st and 99th percentiles in columns (2) and (4). *, **, *** denote significance at the 10%, 5%, and 1% levels.

4.4. Robustness analysis

This section assesses the robustness of the main results to alternative sample definitions, variable constructions, and specification choices.

4.4.1. Alternative activity thresholds

The baseline defines active investors as those executing at least 10 buy orders per month (MIN_BUYS = 10). Using a lower threshold (MIN_BUYS = 5) yields similar results.

The Age coefficient remains positive but decreases to $\beta_{\text{Age}} = 4.46 \times 10^{-5}$, implying a 25-year difference ($25 \times \beta_{\text{Age}}$) increases $IHM_{i,t}^{pw}$ by 0.001115 ($\approx 6.5\%$ of the mean), compared to 9.24×10^{-5} under the baseline.²⁰ This suggests that the relationship between age and herding is stronger among more active investors.

Language is not robust: French dummy becomes small and significant (-0.0009) under MIN_BUYS = 5, but remains insignificant under the baseline specification (MIN_BUYS = 10). Additional checks (interaction model testing whether the sophistication relation differs between French- and Dutch-speaking investors; and investor-level t-test on $\overline{IHM}_i^{pw}$, $p = 0.99$), confirm the absence of meaningful differences (*Figure H6.*).²¹

The negative relationship between herding and performance remains stable, with coefficients of -0.1377 (MIN_BUYS = 5) and -0.2166 (MIN_BUYS = 10) (*see Table 3.*).²² As the result is not driven by activity threshold, baseline threshold is therefore retained as it balances trading activity and sample size.

4.4.2. Alternative eligibility rule (MIN_TRADES)

An alternative definition based on total executed trades (MIN_TRADES = 10) yields the same investor-month sample as the baseline (MIN_BUYS = 10). No observation satisfies the total-trade condition without also meeting the buy-order threshold. As a result, the estimation sample and regression results remain unchanged, and the baseline definition is retained.

²⁰ Obtained by multiplying the estimated Age coefficient ($\beta_{\text{Age}} = 4.46 \times 10^{-5}$) by a 25-year difference: $4.46 \times 10^{-5} \times 25 \approx 0.001115$. Relative to the mean herding level (≈ 0.0171), this corresponds to $\approx 6.5\%$ ($0.001115 / 0.0171$).

²¹ T-test performed on investor-level average herding ($\overline{IHM}_i^{pw}$), comparing French- and Dutch-speaking investors.

²² Coefficients correspond to cross-sectional regressions of performance on average herding intensity ($\overline{IHM}_i^{pw}$). The decrease from -0.1377 to -0.2166 reflects a stronger negative association between herding and performance under the stricter activity threshold.

4.4.3. Winsorization of performance

To verify that the performance results are not driven by extreme values, portfolio performance is winsorized at the 1st and 99th percentiles. The $\overline{IHM}_i^{vw}$ coefficient remains negative and highly significant, decreasing slightly from -0.2166 to -0.2111, while model fit improves marginally. Negative relationship is therefore not driven by extreme observations (*Table II.*).

4.4.4. Equal-weighted IHM

The baseline specification uses the value-weighted herding measure $\overline{IHM}_{i,t}^{vw}$. As a robustness check, the regressions are re-estimated using the equal-weighted measure $\overline{IHM}_{i,t}^{ew}$. Using the equal-weighted specification yields similar results. Coefficient signs are unchanged, and the main relationships are preserved, indicating that findings are not driven by large trades. Although explanatory power is slightly higher, the value-weighted measure is retained as it better reflects capital allocation (*Table I4.*).

4.4.5. Level vs Log specification of trading activity

Trading activity is modeled in levels and in logarithmic form. The log specification improves model fit by approximately 35% in the vw model (0.0188 to 0.0253) and 34% in the ew model (0.0233 to 0.0311), with unchanged signs and significance. This suggests diminishing marginal relationships between trading activity and herding (*Table I5.*).

4.4.6. LSV November sensitivity

November corresponds to the portfolio formation phase, during which most investors are net buyers. As a result, the stock-level proportion of buyers ($p_{j,t}$) is close to the market-wide proportion (p_t), leading to a near-zero dispersion term $|p_{j,t} - p_t|$. Since the $LSV_{j,t}$ measure is based on this dispersion, it is mechanically close to zero in November. Excluding November from the stock-level computation does not affect the investor-level results. Additional details on this mechanical feature are provided in *Appendix E*. This mechanical effect is illustrated in *Figure H11*.

Overall, the robustness checks confirm that the main findings are stable.

5. Discussion

5.1. Interpretation of main findings

Results obtained empirically are useful in providing evidence concerning the presence, determinants, and consequences of herding among retail investors.

First, herding behavior is strongly supported. The Individual Herding Measure is positive on average ($\overline{IHM}^{vw} = 0.0171$), and the null hypothesis of independent trading is clearly rejected ($t = 29.6$, $p < 0.001$). Investors therefore tend to align their trades with aggregate market activity, consistent with prior evidence (Lakonishok et al., 1992; Bikhchandani & Sharma, 2001).

A central finding of this study is the positive relationship between financial sophistication and herding ($\beta = 0.0105$, $p < 0.05$), which contradicts the standard prediction that more informed investors behave more independently. This result highlights a ‘sophistication paradox’: herding is not limited to uninformed investors, but also emerges among relatively sophisticated participants. This challenges the conventional view that sophistication reduces behavioral biases (Merli & Roger, 2013).

There are several mechanisms which may explain this paradox. One reason is that herding may be rational. Sophisticated investors can interpret others’ trades as informative signals when private information is incomplete, leading to convergence in trading behavior (Bikhchandani et al., 1992). Second, the competitive structure of the *Rallye Boursier* encourages imitation of perceived successful strategies in a short-term setting. Third, more sophisticated investors tend to trade more frequently and focus on most popular and followed assets, increasing synchronization across participants. Taken together, these elements suggest that herding may reflect strategic information processing rather than purely behavioral bias.

Age_i is also positively associated with herding, although the relationship remains modest. In this context, age does not appear to be a reliable proxy for investment experience, as participation combines heterogeneous profiles ranging from inexperienced older individuals to financially trained younger participants.

No significant difference is observed between French- and Dutch-speaking investors. While language influences asset selection, it does not affect herding intensity within a shared informational environment. French-speaking investors allocate more to French stocks (14.8%

vs 9.3%; $t = -10.29$, $p < 0.001$), whereas Dutch-speaking investors invest more in Dutch stocks (17.4% vs 14.4%; $t = 6.31$, $p < 0.001$). This is consistent with prior evidence on language-based familiarity biases in portfolio allocation, but does not support the view that linguistic or cultural differences translate into differences in herding behavior. Overall, asset selection and trading synchronization appear to be driven by distinct mechanisms.

Finally, herding is negatively associated with portfolio performance ($\beta_{IHM_i^{pw}} = -0.2166$, $p < 0.001$). Investors with higher levels of IHM_i^{pw} tend to achieve lower returns, suggesting that herding primarily affects timing efficiency. Imitation does not seem to result in faster information aggregation, instead, imitation appears to lead to delayed entry into already-moving positions, consistent with attention-driven trading mechanisms (Barber & Odean, 2008).

At the same time, investor engagement seems to have significant role. More attentive investors tend to achieve higher performance, suggesting that monitoring and discipline are associated with better outcomes than imitation-based strategies.

Thus, it can be stated that herding emerges as a widespread but suboptimal trading pattern, driven more by shared attention and information processing than by superior decision-making.

5.2. Contribution to the literature

In terms of contributing to the existing literature, the thesis makes three main contributions.

First, it provides investor-level evidence on the determinants of herding by documenting a positive relationship between financial sophistication and herding intensity. This finding contrasts with the conventional expectation that more sophisticated investors behave more independently. This ‘sophistication paradox’ suggests that coordinated trading may originate from common interpretations of market signals rather than simple imitation.

Second, this thesis exploits a unique bilingual setting (French- and Dutch-speaking investors) as an underlying cause for herding. However, the absence of robust differences indicates that language alone does not strongly structure herding in this context.

Third, the results highlight the importance of the trading environment. In a short-term and competitive setting such as the *Rallye Boursier*, investors are ranked based on their performance, which creates strong incentives to react quickly to market movements and to follow strategies that appear successful. In this context, herding can become a result of rational

behavior rather than pure behavioral bias. It complements the literature on attention-driven trading by showing that coordinated behavior can emerge even when investors have access to similar information.

In summary, these findings provide insights into how individual characteristics and market context jointly shape herding behavior and its implications for portfolio performance.

5.3. Economic and market implications

The results provide several insights into retail trading behavior and market dynamics.

First, herding appears closely linked to attention-driven mechanisms. The exploratory analyses suggest that herding may be stronger in assets that are frequently traded. When popularity is defined using the top 20% of most traded stocks, the estimated relationship remains positive and statistically significant ($\beta = 0.0248$, $p < 0.001$). Including *PopularShare_{i,t}* in the baseline herding regression also improves model fit, suggesting that exposure to highly traded assets captures an important dimension of investor behavior. This is in line with models where investor direct their attention to well-known or frequently traded securities (Barber & Odean, 2008). Yet, further studies may be needed to identify the exact mechanisms driving herding, for example, attention effects or information diffusion.

Second, findings at the stock-level indicate that herding is driven by coordination rather than directional bias. The *SLSV_{j,t}* distribution is broadly balanced between buy and sell, suggesting that investors move together without systematically pushing prices in one direction.

Third, the negative relationship between herding and performance highlights the imitation cost. Since herding is associated with lower raw returns, investors who deviate from aggregate trading patterns may benefit from better timing. This is consistent with evidence that contrarian behavior can outperform crowd-following strategies in retail markets (Merli & Roger, 2013).

More broadly, these findings suggest that digital and competitive trading environments may amplify collective dynamics. The fact that all traders are interested in the same assets may reflect synchronized behavior rather than independent information processing.

In practical terms, following the crowd is not an effective strategy, even when other investors appear sophisticated. More consistent performance is associated with independent decision-making and active monitoring.

6. Limitations and future research

Several limitations should be considered when interpreting the results.

To begin with, the analysis is conducted within the specific context of the *Rallye Boursier* trading simulation. Participants manage virtual portfolios in a short-term and simulated environment, which may influence trading behavior and investment decisions. Hence, some strategies may be developed due to competition dynamics rather than long-term investment decisions. Although participants react to real market information and real asset prices, they do not directly observe the trades or portfolios of other investors. As such, the observed herding patterns may partly reflect indirect synchronization effects driven by common information, public signals, or shared attention rather than explicit coordination between participants. Moreover, the dataset is based on Belgian investors operating within a relatively small and homogeneous market. Compared to larger or more fragmented markets (e.g. the United States or China), investors are exposed to similar information environments and may be geographically closer, which could reinforce synchronization effects. In addition, the relatively restricted trading universe of 337 traded stocks may concentrate investor attention on a smaller set of assets and create an ‘amplification effect’ for popularity-driven behavior. While this configuration allows for a detailed observation of retail trading at the individual level, caution is required when extrapolating the findings to broader or more diverse financial markets. This reflects a trade-off between internal consistency and external validity.

Afterwards, the empirical analysis is based on a short panel covering only two informative months (December 2019 and January 2020). This limits the ability to study dynamic effects, such as the impact of past performance on subsequent herding.

Then, some variables are only imperfectly observed. Investor engagement is proxied by logins, which capture attention but not fully monitoring intensity or decision quality. Similarly, the sophistication measure combines quiz performance and international diversification, which only partially reflect investor knowledge and behavior. Future research might refine this proxy through the incorporation of other dimensions, such as sectoral diversification. For instance, sophistication could be constructed as a weighted index (e.g. 0.5, 0.25, 0.25), reflecting a simple trade-off between knowledge and diversification. Nevertheless, alternative specifications remain possible, as no unique weighting scheme can fully capture investor sophistication.

Additionally, the herding measurement remains subject to methodological limitations. The LSV measure relies on a binomial benchmark and may underestimate true herding in some settings. In particular, Jurkatis (2022) shows through Monte Carlo simulations that LSV can be interpreted as a test of binomial overdispersion rather than a direct coordination measure. To address this, the analysis combines LSV with IHM (Merli and Roger, 2013). While LSV captures synchronization at the stock level, investor-level IHM evaluates whether investors trade in the same direction as the crowd, allowing a more direct assessment of heterogeneity.

Furthermore, there is an emphasis on identifying robust empirical relationships rather than establishing causal effects. These links, including the negative association between herding and performance, and the positive relationship between sophistication and herding might have roots in other variables not directly observed. Establishing causality would require additional identification strategies beyond the scope of this dataset. This paper's intention is therefore not to provide a predictive or causal model, but to document and interpret behavioral patterns, which represents a standard and necessary first step in empirical research.

Relatedly, the econometric specification is intentionally parsimonious. While this facilitates interpretation, additional controls could be incorporated in future work, such as index membership, macroeconomic indicators (e.g. inflation or economic sentiment indices), financial variables (e.g. volatility indices such as the VIX), or attention proxies (e.g. Google Trends or social media activity indicators). Institutional factors, international events, and calendar effects could also influence trading.

Despite these limitations, the dataset provides detailed transaction-level information, allowing a direct analysis of coordinated trading behavior in a controlled environment.

Future research could extend this analysis in several directions. Longer time horizons would allow the study of dynamic feedback effects between performance and herding. Additional investor characteristics (e.g. gender or wealth) could refine behavioral heterogeneity. Risk-adjusted performance measures construction would bring additional insight into portfolio efficiency. In the end, linking trading activity to external information sources such as news or social media would add some value to the understanding of attention-driven mechanisms.

7. Conclusion

This thesis examines the presence, determinants, and consequences of herding behavior among retail investors participating in the *Rallye Boursier* trading simulation. It addresses the following research question: *How can retail herding be robustly measured and explained through the combined influence of socio-demographic, cultural, and market participation factors, and what implications does it have for portfolio performance?*

The results provide a clear answer. Herding is present, measurable, and economically meaningful. The average individual herding measure is positive ($\overline{IHM}^{vw} \approx 0.0171$), indicating systematic alignment between individual trades and aggregate market activity. Using a dual approach combining the Lakonishok, Shleifer and Vishny measure (LSV) and the Individual Herding Measure (IHM), the analysis shows that investors systematically align their trades with aggregate market activity. This pattern is robust across specifications, activity thresholds, and alternative constructions of the herding measure, and is supported by extensive diagnostic checks as well as complementary evidence reported in the appendices. Altogether, these findings confirm that coordinated trading is a central feature of individual investor behavior.

Beyond its existence, herding is related to both individual characteristics and market dynamics. A key contribution of this thesis is the identification of a ‘sophistication paradox’: more sophisticated investors tend to herd more, not less. This challenges the conventional view that financial knowledge reduces behavioral biases. Instead, it suggests that coordinated behavior may arise from common interpretations of market signals, shared attention, and strategic adaptation in a competitive environment. Age is also positively related to herding, although modestly, while linguistic differences are associated with asset selection patterns, they do not affect the intensity of imitation once investors operate within a shared informational context.

Moreover, herding is associated with lower raw portfolio performance. The estimated coefficient on average herding intensity is economically meaningful ($\beta \approx -0.2166$), indicating that investors who more closely follow aggregate trading flows tend to achieve lower returns. This suggests that imitation primarily affects timing rather than improving information aggregation. The result is consistent with attention-driven trading mechanisms identified by Barber and Odean (2001; 2008), who found that investors focus on popular assets and trade after price movements have already occurred. Additional evidence on popular stocks and stock-level coordination supports this interpretation, as herding is due to synchronized reaction to

common signals rather than persistent directional bias.

Taken together, these findings contribute to the literature in three main ways. First, they provide investor-level evidence by linking stock-level (LSV) and individual-level (IHM) measures within a unified framework. Second, they document the ‘sophistication paradox’, showing that coordinated behavior is not limited to uninformed investors. Third, they highlight the role of trading environment, where shared information contributes to synchronized behavior.

There are also broader implications for financial markets in today’s environment. Given the increasing digitalization, where information circulates rapidly, particularly through social networks, investors may converge toward similar decisions without direct coordination. Episodes such as the GameStop short squeeze illustrate how collective attention can translate into rapid market movements, even in the absence of centralized information processing.

These conclusions should nevertheless be interpreted within the limits of the study. The analysis relies on a short-term simulation and observational data, which does not allow for causal interpretation. While the results are robust and internally consistent, extending the analysis to longer time horizons and richer datasets would help better understand these mechanisms. Future research could further investigate whether stock popularity and investor attention amplify synchronization effects in retail trading environments.

Overall, this thesis shows that retail investors do not act independently, and that even relatively sophisticated participants may converge toward similar trading decisions.

Herding does not necessarily reflect irrationality, but even sophisticated investors may move together in ways that are individually costly and collectively inefficient.

8. Appendices

8.A. Appendix A. LSV adjustment factor computation

The LSV measure requires the adjustment factor: $AF_{j,t} = \mathbb{E} \left[\left| \frac{X_{j,t}}{I_{j,t}} - p_t \right| \right]$, where $X_{j,t}$ follows a binomial distribution with parameters $I_{j,t}$ and p_t . This factor corrects for the expected deviation in the proportion of buyers under the null hypothesis of independent trading.

In practice, two computation methods are implemented. The adjustment factor is computed exactly from the binomial distribution when the number of active investors is moderate, and approximated using a normal formula when it becomes sufficiently large. The rule is:

$$AF_{j,t} = \begin{cases} \text{Exact binomial} & \text{if } I_{j,t} \leq 2000 \\ \text{Normal approximation} & \text{if } I_{j,t} > 2000 \end{cases}$$

with $n_{max \text{ exact}} = 2000$ in the baseline specification.

Importantly, this rule is **exclusive**: for each stock-month, the adjustment factor is computed either using the exact binomial formula or the normal approximation, but never a combination of both. For example, if $I_{j,t} = 2500$, the normal approximation is applied to the entire observation.

The sensitivity of this choice is also studied through alternative thresholds of 1000 and 3000. Nonetheless, within the selected sample, the maximum number of active investors per stock month is 534. As a result, all observations satisfy $I_{j,t} \leq 1000$, and the exact binomial computation is always used.²³ This approach ensures efficiency in computation and exactness when sample sizes are moderate.

Thus, varying the threshold from 1000 to either 2000 or 3000 leads to the same adjustment factors, herding measures, and regression results. Consequently, it can be stated with certainty that the normal approximation, although implemented in the code for completeness, is never triggered in the final sample.

²³ Throughout the thesis, AF remains defined under the binomial null of independent trading. The normal formula is used only as a computational approximation for large $I_{j,t}$; it does not replace the theoretical definition of the adjustment factor.

8.B. Appendix B. Equity universe filter

The trading panel is restricted to a clean equity universe. Only securities satisfying $TYPE = STOCK$ and $STOCK_{TYPE} \in \{\text{share, trust}\}$ are retained.

This filter removes non-equity instruments that may appear in the raw orders dataset. The goal is to compute herding measures on a homogeneous set of common stocks, ensuring comparability across investors and stock-month observations.

The filter is applied before constructing the eligibility sample. The procedure therefore consists of restricting the universe to eligible equity securities, keeping executed trades only, applying the investor-month activity threshold, and computing stock-level herding (stock-month LSV) and investor-level (investor-month IHM). Applying the filter first prevents activity thresholds from being influenced by non-equity trades and ensures a consistent security universe.

In practice, the restriction does not change the final estimation sample. After filtering, no remaining ISIN lies outside the equity universe. The initial panel contains 2,697 investor-month observations corresponding to 2,211 investors. After removing observations with missing values in regression variables, the final estimation sample contains 2,693 investor-month observations and 2,208 unique investors, observed in December and January.

The filter thus improves dataset consistency without affecting sample size or regression results, ensuring that herding measures consistently reflect trading in common equities only.

8.C. Appendix C. Numerical illustration of LSV, SLSV & IHM

This appendix provides a simplified numerical example to illustrate the construction of the LSV, SLSV and IHM measures. The example is based on three assets (A, B, C) observed over three months ($t = 1, 2, 3$).

8.C.1. Stock-level trading activity

Table C1. reports the number of buyers and sellers for each stock and month.

Table C1. Trading activity by stock and month

Month t	Stock j	Net buyers ($B_{j,t}$)	Net Sellers ($S_{j,t}$)	Number of active investors ($I_{j,t} = B_{j,t} + S_{j,t}$)	Purchase intensity ($p_{j,t} = \frac{B_{j,t}}{I_{j,t}}$)
t ₁	A	8	2	10	0.80
t ₁	B	5	5	10	0.50
t ₁	C	2	8	10	0.20
t ₂	A	6	4	10	0.60
t ₂	B	3	7	10	0.30
t ₂	C	7	3	10	0.70
t ₃	A	4	6	10	0.40
t ₃	B	6	4	10	0.60
t ₃	C	5	5	10	0.50

Note : $p_{j,t}$ represents the proportion of investors trading stock j who are buyers.

The market-wide proportion of buyers (p_t) is computed as the ratio of total buyers to total trades across all stocks:

$$p_t = \frac{\sum_j B_{j,t}}{\sum_j (B_{j,t} + S_{j,t})}$$

$$t_1: p_t = (8 + 5 + 2) / (10 + 10 + 10) = 0.50$$

$$t_2: p_t = (6 + 3 + 7) / (10 + 10 + 10) = 0.53$$

$$t_3: p_t = (4 + 6 + 5) / (10 + 10 + 10) = 0.50$$

Unlike $p_{j,t}$, which is stock-specific, p_t is computed at the market level. It therefore does not depend on j and captures the overall tendency to buy in the market during period t .

8.C.2. Computation of LSV and SLSV

The LSV measure captures the extent to which trading activity on a given stock deviates from the overall market behavior. Formally, the LSV measure is defined as: $LSV_{j,t} = |p_{j,t} - p_t| - AF_{j,t}$ where $AF_{j,t}$ is an adjustment factor correcting for random trading.

In this simplified example, $AF_{j,t} = 0$ to focus on the intuition of the measure. The LSV therefore simplifies to: $LSV_{j,t} = |p_{j,t} - p_t|$

The signed version of the measure is defined as:

$$SLSV_{j,t} = \begin{cases} LSV_{j,t} & \text{if } p_{j,t} > p_t \text{ (excess buying; positive values)} \\ -LSV_{j,t} & \text{if } p_{j,t} < p_t \text{ (excess selling; negative values)} \\ 0 & \text{if } p_{j,t} = p_t \text{ (neutral cases; zero)} \end{cases}$$

Table C2. LSV and SLSV

Month t	Stock j	Purchase intensity $(p_{j,t} = \frac{B_{j,t}}{I_{j,t}})$	Market-wide proportion of buyers p_t	LSV _{j,t}	SLSV _{j,t}
t1	A	0.80	0.50	0.30	+0.30
t1	B	0.50	0.50	0.00	0.00
t1	C	0.20	0.50	0.30	-0.30
t2	A	0.60	0.53	0.07	+0.07
t2	B	0.30	0.53	0.23	-0.23
t2	C	0.70	0.53	0.17	+0.17
t3	A	0.40	0.50	0.10	-0.10
t3	B	0.60	0.50	0.10	+0.10
t3	C	0.50	0.50	0.00	0.00

Note: These values show that herding direction varies across assets and over time.

8.C.3. Investor-level herding (IHM)

We now consider a representative investor i .

The $IHM_{i,t}^{vw}$ for investor i during period t is defined as a value-weighted average of the signed stock-level signals:

$$IHM_{i,t}^{vw} = \frac{\sum_{j=1}^J (\Delta n_{i,j,t} P_{j,t}) SLSV_{j,t}}{\sum_{j=1}^J |\Delta n_{i,j,t}| P_{j,t}}$$

where:

- $\Delta n_{i,j,t} = n_{i,j,t} - n_{i,j,t-1}$ is the change in the number of shares held by investor i in stock j between $t - 1$ and t or, in other words, the net change in shares (buys minus sells) of stock j held by investor i in month t ;
- $P_{j,t}$ is the average price of stock j over the period $[t - 1; t]$;
- The denominator corresponds to the total monetary value of transactions executed by investor i during the period (in €).

Table C3. Investor trades (month t2)

Stock j	Net change in shares (buys minus sells) of stock j held by investor i in month t $\Delta n_{i,j,t}$	Average stock price $P_{j,t}$ (€)	$SLSV_{j,t}$
A	+10	50	+0.07
B	-5	100	-0.23
C	+8	25	+0.17

***Note:** Investor-level trades are illustrative and are not directly derived from the aggregated data in Table C1. They are constructed to provide a simple and intuitive example of how stock-level herding signals (SLSV) are translated into an individual herding measure (IHM).*

8.C.3.1. Step 1 - Numerator computation

A: $(+10) * 50 * 0.07 = 35$

B: $(-5) * 100 * (-0.23) = 115$

$$C: (+8) * 25 * 0.17 = 34$$

$$\text{Numerator} = 35 + 115 + 34 = 184$$

8.C.3.2. Step 2 - Denominator computation

$$A: |(+10)| * 50 = 500$$

$$B: |(-5)| * 100 = 500$$

$$C: |(+8)| * 25 = 200$$

$$\text{Denominator} = 1200$$

8.C.3.3. Step 3 - IHM computation in t_2

$$\text{IHM}_{i,t_2} = 184 / 1200 = 0.1533$$

8.C.4. Interpretation

8.C.4.1. Interpretation of LSV and SLSV

The $\text{LSV}_{j,t}$ measure captures the extent to which trading on a given stock deviates from the overall market behavior. A higher value of $\text{LSV}_{j,t}$ indicates stronger synchronization among investors on that stock.

The signed version ($\text{SLSV}_{j,t}$) adds direction to this measure. A positive value indicates that investors disproportionately buy the stock compared to the market average (excess buying), while a negative value reflects excess selling.

For instance, as seen in *Table C2.*, in period t_l , stock A exhibits a high positive $\text{SLSV}_{j,t}$ (+0.30), indicating strong buying pressure relative to the market. In contrast, stock C shows a negative $\text{SLSV}_{j,t}$ (-0.30), reflecting strong selling pressure. Stock B, with $\text{SLSV}_{j,t}$ equal to zero, follows the market average and does not exhibit herding.

Overall, these results illustrate that herding is not uniform across assets and varies both in intensity and direction over time.

8.C.4.2. Interpretation of $IHM_{i,t}^{vw}$

The computed value of $IHM_{i,2}^{vw}$ is equal to 0.1533, which is strictly positive. This indicates that the investor's trading decisions are aligned with the prevailing market behavior.

More precisely, the investor buys stocks that are subject to excess buying pressure (positive $SLSV_{j,t}$) and sells stocks that are subject to excess selling pressure (negative $SLSV_{j,t}$). This consistency between individual trades and market-wide signals results in a positive herding measure.

From an economic perspective, the $IHM_{i,t}^{vw}$ captures whether an investor tends to follow the crowd or act in opposition to it. A positive value reflects herding, while a negative value would indicate contrarian behavior.

This example therefore illustrates how stock-level coordination ($LSV_{j,t}$ and $SLSV_{j,t}$) is translated into an investor-level measure, and how individual trading decisions can be evaluated relative to aggregate market dynamics.

Importantly, the magnitude of $IHM_{i,t}^{vw}$ depends not only on the direction of trades, but also on their monetary importance, as larger trades receive a higher weight in the value-weighted specification.

8.C.4.3. $IHM_{i,t}^{ew}$ as a robustness check

As a robustness check, the investor-level herding measure can also be computed using an equal-weighted approach. In this case, each traded stock contributes equally to the measure, regardless of the monetary size of the transaction.

The $IHM_{i,t}^{ew}$ specification is defined as: $IHM_{i,t}^{ew} = \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \text{sign}(\Delta n_{i,j,t}) * SLSV_{j,t}$ where $N_{i,t}$ is the number of traded stocks.

This formulation can be interpreted as a simple average of stock-level herding signals, weighted only by the direction of the investor's trades. For each stock j , the term $\text{sign}(\Delta n_{i,j,t})$ takes the value +1 if the investor is a net buyer and -1 if the investor is a net seller.

The product $\text{sign}(\Delta n_{i,j,t}) * SLSV_{j,t}$ therefore captures whether the investor trades in the same direction as the market: it is positive when the investor follows the prevailing market pressure

(buying when $SLSV_{j,t} > 0$ or selling when $SLSV_{j,t} < 0$); and negative when the investor trades against it. By averaging these contributions across all traded stocks, the $IHM_{i,t}^{ew}$ measures the overall directional alignment of the investor with the market, without giving more weight to larger positions. Using the same example as above (period t_2):

- Stock A: buy $\rightarrow$ sign = +1, SLSV = +0.07 $\rightarrow$ contribution = (+1 * 0.07) = +0.07

- Stock B: sell $\rightarrow$ sign = -1, SLSV = -0.23 $\rightarrow$ contribution = (-1 * -0.23) = +0.23

- Stock C: buy $\rightarrow$ sign = +1, SLSV = +0.17 $\rightarrow$ contribution = (+1 * 0.17) = +0.17

IHM_{i,t_2}^{ew} equals therefore: $(0.07 + 0.23 + 0.17) / 3 = 0.1567$

$IHM_{i,t}^{ew}$ is > 0 , indicating that the investor's trades are aligned with the prevailing market direction.

In this example, buying stocks under excess buying pressure and selling stocks under excess selling pressure both contribute positively to the measure. This reflects behavior consistent with the crowd. Compared to the value-weighted measure, the equal-weighted specification gives equal importance to all trades, regardless of their size. It therefore captures directional alignment only, while the vw measure also reflects the economic magnitude of positions.

Importantly, the equally weighted measure is not mechanically positive. It becomes negative when the investor trades against the prevailing market direction, i.e., buying stocks under selling pressure or selling stocks under buying pressure.

Table C4. $IHM_{i,t}^{ew}$ is not always positive

SLSV (market signal)	Investor decision	Contribution to $IHM_{i,t}^{ew}$	Interpretation
+	BUY (+)	(+) * (+) = (+)	Herding
+	SELL (-)	(+) * (-) = (-)	Contrarian herding
-	SELL (-)	(-) * (-) = (+)	Herding
-	BUY (+)	(-) * (+) = (-)	Contrarian herding

Note: A positive contribution indicates that the investor trades in the same direction as the market, while a negative contribution reflects trading against the prevailing market pressure.

8.D. Appendix D. Model fit interpretation and literature comparison

This subsection explains how model fit should be interpreted in individual-level behavioral data and situates the results relative to the literature.

8.D.1. Why low R^2 is expected and acceptable

The determinants regression yields $R^2 \approx 0.025$ and Within $R^2 \approx 0.023$ in the vw specification. Although these values may appear small, they are consistent with data structure.

$IHM_{i,t}^{vw}$ reflects discrete trading decisions by investors, which depend on many unobserved factors such as private information, liquidity needs, attention constraints, portfolio rebalancing, and psychological biases. If the data-generating process is written as $IHM_{i,t}^{vw} = X_{i,t} \beta + \varepsilon_{i,t}$, the residual term $\varepsilon_{i,t}$ captures substantial idiosyncratic variation. Even when the systematic component $X_{i,t} \beta$ is economically meaningful, the unexplained variance remains large.

Observable characteristics such as Age and Sophistication therefore explain only a small share of total variation. For example, the Age coefficient ($9.24 * 10^{-5}$) is statistically significant but economically modest.

In this context, the objective is structural inference rather than prediction. Interpretation focuses on the sign, magnitude, and significance of coefficients, while R^2 plays a secondary role. Low explanatory power is therefore expected when modeling noisy individual trading behavior.

8.D.2. Comparison with Merli & Roger (2013)

Merli and Roger (2013) report $R^2 \approx 0.25$ in their study of mutual fund herding, whereas the present analysis finds R^2 between 0.02 and 0.03 for individual investors.

These values are not directly comparable for three reasons.

First, the unit of observation differs. Merli and Roger (2013) use fund-level cross-sectional data, while this study relies on an investor-month panel with 2,693 observations across 2,208 investors. Fund-level aggregation smooths idiosyncratic variation and mechanically increases explanatory power by construction.

Second, the dependent variable differs. Fund herding reflects coordinated portfolio adjustments by professional managers and is typically more persistent. In contrast, $IHM_{i,t}^{vw}$ captures

individual trading relative to the market benchmark and varies strongly at the micro level.

Third, the sample composition differs. Professional managers operate under institutional constraints and investment guidelines that reduce behavioral dispersion, whereas investors display greater heterogeneity and noise.

Low R^2 values are therefore common in individual-level behavioral finance. Despite the lower explanatory power, both studies identify statistically significant determinants of herding. In this analysis, the positive effect of Sophistication and the negative relationship between herding and performance remain highly significant.

8.D.3. Within R^2 interpretation in FE models

The determinants regressions include month fixed effects: $IHM_{i,t} = X_{i,t}\beta + \lambda_t + \varepsilon_{i,t}$. The Within R^2 measures the share of variation in $IHM_{i,t}$ explained by $X_{i,t}$ after removing the month effects λ_t . In the baseline specification, Within $R^2 \approx 0.0229$ in the value-weighted model and 0.0276 in the equal-weighted model.

Consider an alternative specification including investor FE in addition to month effects. In that case, the total R^2 rises mechanically to about 0.56, because individual dummies absorb time-invariant heterogeneity. However, the Within R^2 falls close to zero.

This occurs because investor FE absorb all time-invariant variables, including Age, Language, and Sophistication, making their coefficients unidentified. The higher total R^2 therefore reflects absorbed heterogeneity rather than the explanatory power of the covariates.

For this reason, the preferred specification includes month FE only. This controls for common time shocks while allowing the estimation of time-invariant investor characteristics. The Within R^2 is therefore the appropriate measure of model fit.

Overall, modest R^2 values are consistent with behavioral data and do not diminish the statistical or economic relevance of the estimated relationships.

8.E. Appendix E. Sensitivity to LSV computation

8.E.1. November 2019: Mechanical LSV equal to zero

Diagnostic checks show that in November 2019, the stock-level proportion of buyers equals the market-wide benchmark: $p_{j,t} = p_t = 1$. This can be explained step by step.

First of all, $p_{j,t}$ denotes the proportion of active investors who are net buyers of stock j in month t . In November 2019, this proportion is equal to 1 for the stocks under consideration. In other words, for a given stock month, all active investors are on the buy side, and none are net sellers.

Then, p_t denotes the market-wide proportion of buyers in month t , computed across all considered stock-month observations. In November 2019, this benchmark is also equal to 1. This means that buying is not only dominant for a specific stock, but generalized across the whole market at that time. Formally, this implies: $p_{j,t} - p_t = 1 - 1 = 0$

The $LSV_{j,t}$ measure is defined as: $LSV_{j,t} = |p_{j,t} - p_t| - AF_{j,t}$

The first term is therefore: $|p_{j,t} - p_t| = |1 - 1| = 0$

The adjustment factor $AF_{j,t}$ is also equal to zero in this case. Intuitively, this factor corrects for the amount of dispersion that could arise mechanically from random variation in buy and sell decisions under the null of independent trading. Here, however, there is no such variation to correct for: all investors behave in the same direction, and the proportion of buyers is at its upper boundary value.

Formally, the adjustment factor is defined as: $AF_{j,t} = \mathbb{E} \left[\left| \frac{X}{I_{j,t}} - p_t \right| \right], X \sim \text{Bin}(I_{j,t}, p_t)$

Equivalently, using the notation introduced earlier, $X = B_{j,t}$, where $B_{j,t}$ denotes the number of buyers for stock j in month t . In November 2019, $p_t = 1$, so the binomial distribution becomes degenerate: all investors are buyers with probability of one.

Hence, $X = B_{j,t} = I_{j,t}$ with probability of one, which implies deterministically: $\frac{X}{I_{j,t}} = \frac{B_{j,t}}{I_{j,t}} = 1$

Therefore, $AF_{j,t} = \mathbb{E} \left[\left| \frac{X (=B_{j,t})}{n (=I_{j,t})} - p_t \right| \right] = \mathbb{E}[|1 - 1|] = 0$

The same conclusion follows under the normal approximation used for large $I_{j,t}$.

In that case, $AF_{j,t} \approx \sqrt{\frac{2}{\pi}} \sqrt{\frac{p_t(1-p_t)}{I_{j,t}}}$ and when $p_t = 1$, $AF_{j,t} \approx \sqrt{\frac{2}{\pi}} \sqrt{\frac{1(1-1)}{I_{j,t}}} = 0$

Thus, both the exact binomial computation and the normal approximation lead to the same result in this specific case. The LSV measure therefore simplifies to: $LSV_{j,t} = |1 - 1| - 0 = 0$

This result is mechanical rather than behavioral. November corresponds to the portfolio formation phase of the competition, during which investors were overwhelmingly net buyers.

Because most participants were building positions rather than selling or rebalancing, cross-sectional variation in trading direction was virtually absent. In such a setting, $LSV_{j,t}$ cannot identify meaningful stock-level differences in herding, since all stock-month observations mechanically collapse to the same benchmark outcome. This is why November is not informative for the stock-level herding analysis.

8.E.2. Focus on December and January

To avoid this mechanical constraint, the regression panel is restricted to December 2019 and January 2020. Across these months, 575 out of 671 stock-month observations are measurable (85.7%), while 14.3% are excluded because fewer than 10 investors traded the stock.

Cross-sectional variation in herding is substantial. Average $LSV_{j,t} = 0.0299$ in December and 0.0405 in January, with SD of 0.0707 and 0.0837. $SLSV_{j,t}$ is slightly negative (-0.0066 in December and -0.0082 in January) with roughly balanced positive and negative realizations.

8.E.3. Impact on final results

Since November is excluded, the mechanical pattern does not affect the empirical results. The investor-month panel contains 2,693 observations across 2,208 investors, including 1,859 observations in December and 838 in January.

The positive effect of Sophistication, the modest Age effect, and the negative and significant relationship between herding and performance are therefore driven entirely by months with genuine cross-sectional variation in $LSV_{j,t}$.

Overall, the November anomaly reflects a structural feature of the portfolio formation phase, not an issue with the $LSV_{j,t}$ computation. Once excluded, the herding measures behave as expected and the main results remain unchanged.

8.F. Appendix F. Sample construction & descriptive statistics

8.F.1. Sample construction & activity threshold

The empirical analysis is conducted at the investor-month level. The final dataset contains 2,693 investor-month observations corresponding to 2,208 unique investors. Because the sample is restricted to months in which investors execute at least ten buy orders, some investors appear only once in the panel while others appear multiple times (up to two times). This is typical in short trading competitions where participation intensity varies across months.

The activity threshold is applied at the monthly level to ensure that the $IHM_{i,t}$ is computed only when investors display sufficient trading activity. Applying the threshold over the entire sample period would retain investors who are active in only one month but inactive in others, leading to unreliable herding estimates.

This restriction focuses the analysis on relatively active investors and excludes those with very limited trading activity. While this may slightly bias the sample toward more active participants, the filter is necessary to obtain reliable herding measures. Moreover, robustness tests using alternative thresholds (e.g., a minimum of five buy orders) yield similar results, indicating that the main findings are not driven by the cutoff. This may limit the representativeness of the sample for less active participants.

8.F.2. International diversification & geographical concentration

The number of distinct countries traded per investor shows meaningful cross-sectional variation. On average, investors trade in 5.55 countries (median of 5), with values ranging from 1 to 12. Most investors operate in 4 to 7 countries, indicating a moderate degree of international diversification, while purely domestic portfolios remain rare (8 out of 2,208 investors).

At the aggregate level, trading activity is highly concentrated. The top three countries (Belgium, the United States, and the Netherlands) account for 76.2% of total trades, and the top five for 90.3%, despite the presence of 13 distinct countries in the dataset.

Trading is also predominantly European, with 71.4% of transactions occurring within EU countries. This partly reflects the composition of the investment universe, but also suggests a preference for geographically familiar markets. At the same time, some degree of international exposure persists, particularly towards large markets such as the United States.

8.G. Appendix G. Additional empirical evidence

8.G.1. Coordination effects

To examine whether herding increases when more investors trade the same asset, the relationship between the stock-level herding measure and the number of active investors is analyzed. The variable $I_{j,t}$ denotes the number of investors trading stock j in month t .

A simple correlation shows a positive association between investor participation and herding intensity ($\rho = 0.124$). This pattern is confirmed in a regression: $LSV_{j,t} = \alpha + \beta I_{j,t} + \varepsilon$, where the estimated coefficient is positive and statistically significant ($\beta = 0.0001$, $p < 0.01$).

Because the number of investors is highly skewed, a logarithmic specification provides a more interpretable model: $LSV_{j,t} = \alpha + \beta \log(I_{j,t}) + \varepsilon$. The coefficient on $\log(I_{j,t})$ is positive and highly significant ($\beta = 0.013$, $p < 0.001$). A doubling of the number of active investors is associated with an increase of approximately 0.009 in the LSV measure, corresponding to roughly 25% of the average herding intensity. Results remain similar when controlling for month fixed effects.

8.G.2. Popularity and herding intensity (IHM)

Herding appears stronger on highly traded assets. Popular stocks are defined each month according to their number of trades, allowing stock popularity to vary over time. A paired t-test comparing investor-level herding on popular versus less actively traded stocks shows that the difference is positive and statistically significant for the top 10% and top 20% thresholds, but not for the broader top 30% definition. This suggests that the effect is concentrated among the most actively traded stocks rather than spread across all frequently traded assets.

To further examine this relationship, a variable $PopularShare_{i,t}$ is constructed as the share of the total trading value of investor i in month t allocated to stocks classified as popular. The following regression model is estimated:

$$IHM_{i,t}^{vw} = \beta_0 + \beta_1 PopularShare_{i,t} + \beta_2 Sophistication_i + \beta_3 Age_i + \beta_4 French_i \\ + \beta_5 \log(Trades_{i,t}) + \lambda_t + \varepsilon_{i,t}$$

where $PopularShare_{i,t}$ is computed according to the popularity threshold considered. *Table G2* reports both the paired t-test results and the corresponding regression estimates across alternative definitions of popularity.

Table G2 Herding intensity ($IHM_{i,t}^{vw}$) and popular stocks

Popular stock definition (most traded stocks)	Paired t-test		Regression results	
	Paired t-test mean difference $\overline{IHM}_{pop}^{vw} - \overline{IHM}_{other}^{vw}$	p-value	$PopularShare_{i,t}$ coefficient (SE)	Within R ²
Top 10%	0.00399	0.0002	0.0261*** (0.0035)	0.05011
Top 20%	0.00323	0.0035	0.0248*** (0.0032)	0.04713
Top 30%	0.00058	0.653	0.0171*** (0.0035)	0.03204

Notes: Popular stocks are defined each month according to their number of trades, allowing stock popularity to vary over time rather than being fixed over the full sample period. Columns (2) and (3) report paired t-tests comparing investor-level herding on popular and non-popular stocks. Columns (4) and (5) report coefficients from regressions where $IHM_{i,t}^{vw}$ is regressed on $PopularShare_{i,t}$ and control variables ($Sophistication_i$, Age_i , $French_i$, $\log(Trades_{i,t})$), including month FE, with SE clustered at the investor level. The 20% threshold is retained as the baseline specification because it provides a balance between statistical strength and representativeness across both the paired t-test and regression approaches. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table G2. shows that herding is stronger on the most actively traded stocks. The paired t-test is significant for the top 10% and top 20% thresholds, but not for the top 30% definition, suggesting that the effect becomes diluted when the popularity group is defined too broadly.

Across all thresholds, the $PopularShare_{i,t}$ coefficient remains positive and highly significant. Investors allocating a larger share of their trading value to popular stocks therefore exhibit stronger herding behavior, even after controlling for individual characteristics, trading activity, and month fixed effects. The relatively higher Within R² values suggest that stock popularity captures an additional behavioral dimension beyond standard investor characteristics.

These results indicate that the positive relationship between stock popularity and herding intensity holds across multiple definitions of popularity, though it is most pronounced for narrower definitions of popularity.

8.G.3. Stock popularity and stock-level herding (LSV and SLSV)

To complement the investor-level analysis, stock-level herding measures are examined using the $LSV_{j,t}$ and $SLSV_{j,t}$ metrics. Stock popularity is positively associated with herding intensity across all thresholds, although the relationship becomes weaker when the popularity definition is more inclusive. The coefficient on the Popular dummy is positive and highly significant at the top 10% threshold ($\beta = 0.0377$, $p < 0.001$), remains significant at 20% ($\beta = 0.0196$, $p = 0.010$), and becomes only marginally significant at 30% ($\beta = 0.0130$, $p = 0.058$).²⁴

In contrast, no consistent relationship is found for directional herding. The coefficient on $|SLSV_{j,t}|$ is not statistically significant at the 10% and 20% thresholds. The negative coefficient observed at the 30% threshold likely reflects the dilution of the popular group rather than a genuine directional reversal. Overall, stock popularity affects the magnitude of coordinated trading rather than its direction. Results are summarized in *Table G3*.

Table G3. Stock-level herding (LSV and SLSV) and popular stocks

Popular stock definition	Mean $LSV_{j,t}$ (popular)	Mean $LSV_{j,t}$ (other) ²⁵	β ($LSV_{j,t}$)	p-value	β ($ SLSV_{j,t} $)	p-value
Top 10%	0.0663	0.0305	0.0377***	<0.001	0.0106	0.476
Top 20%	0.0480	0.0307	0.0196**	0.010	-0.0051	0.420
Top 30%	0.0414	0.0311	0.0130*	0.058	-0.0120	0.655

Notes: Popular stocks are defined each month according to their number of trades.

The 20% threshold is retained as the baseline specification, provides a balance between statistical strength and consistency. The table reports stock-level regressions of herding intensity ($LSV_{j,t}$) and absolute signed herding ($|SLSV_{j,t}|$) on a Popular dummy, with month fixed effects.

*, **, *** denote significance at the 10%, 5%, and 1% levels.

²⁴ Robustness checks using alternative thresholds confirm that extreme definitions lead to less stable results. At 1%, the popular group contains only 7 stock-month observations, resulting in imprecise estimates ($\beta = 0.0058$, $p = 0.486$), and the paired t-test yields a negative difference. At 50%, the coefficient remains positive and significant ($\beta = 0.0111$, $p = 0.025$), but the paired t-test is no longer significant ($p = 0.848$).

²⁵ Mean LSV values compare average herding intensity between popular and non-popular stocks. The difference is consistent with the estimated coefficient on the Popular dummy (e.g., 0.0196 at the 20% threshold).

8.G.4. Most herded stocks by month

Table G4. provides a stock-level illustration of herding by identifying the most extreme buy- and sell-side cases in each month. The composition of these stocks varies across periods, indicating that herding is not concentrated on a fixed set of securities but evolves over time.

Strongest herding episodes are concentrated in a limited number of stock-months, suggesting that synchronization arises in specific situations rather than uniformly across the market.

The table is descriptive and does not rely on statistical testing. Some observations are based on a limited number of active investors, which may affect the stability of the estimates. Results should therefore be interpreted as illustrative evidence of stock-level patterns. The number of active investors ($I_{j,t}$) is reported to indicate the underlying participation.

Table G4. Top 3 buy-side & sell-side herding stocks by month

Month t	Herding side	Rank	Stock name	ISIN	$LSV_{j,t}$ (%)	$ SLSV_{j,t} $ (%)	Number of active investors $I_{j,t}$
Dec	Buy	1	Caterpillar	US1491231015	20.3	20.3	19
Dec	Buy	2	Newmont Goldcorp	US6516391066	18.4	18.4	249
Dec	Buy	3	Sipef	BE0003898187	18.1	18.1	187
Dec	Sell	1	Iliad	FR0004035913	40.5	40.5	18
Dec	Sell	2	TKH Group	NL0000852523	30.8	30.8	16
Dec	Sell	3	Greenyard	BE0003765790	24.7	24.7	211
Jan	Buy	1	Sofina	BE0003717312	25.8	25.8	81
Jan	Buy	2	Immobel	BE0003599108	24.8	24.8	29
Jan	Buy	3	Bolloré	FR0000039299	20.1	20.1	16
Jan	Sell	1	Walt Disney	US2546871060	26.2	26.2	54
Jan	Sell	2	Synergie	FR0000032658	25.6	25.6	14
Jan	Sell	3	OCI	NL0010558797	23.9	23.9	27

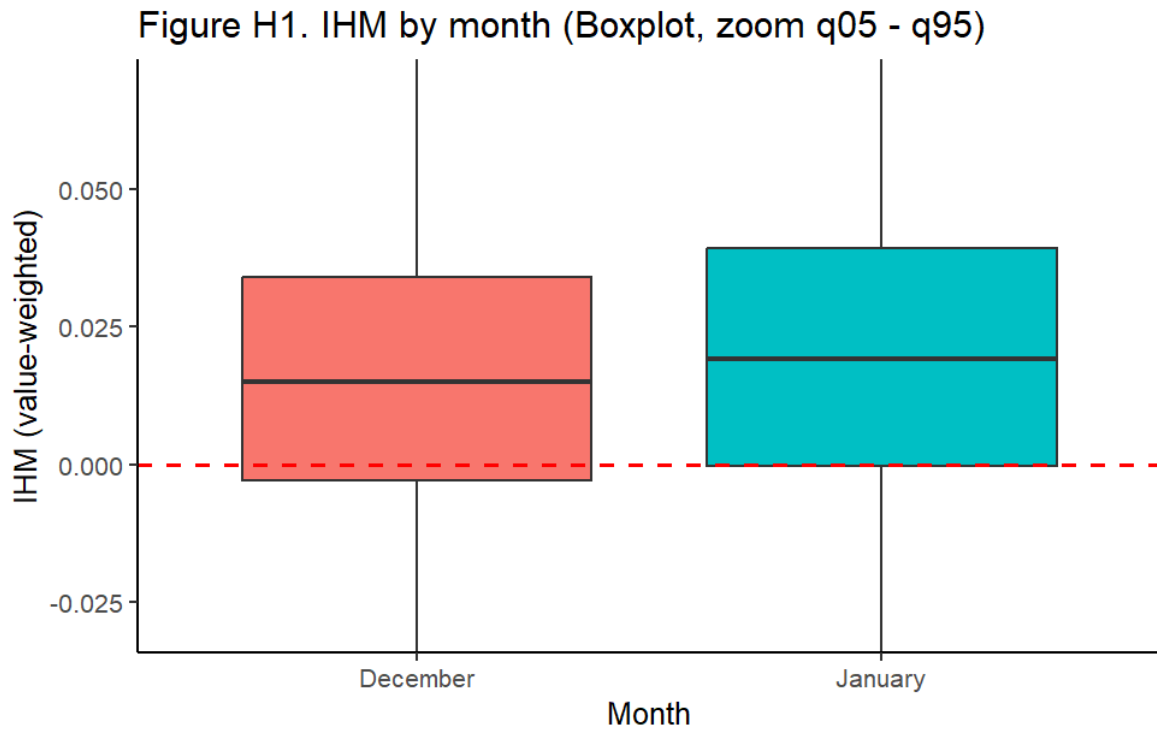
Notes: Stocks are ranked by the absolute value of signed herding ($|SLSV_{j,t}|$), separately for buy-side herding ($SLSV_{j,t} > 0$) and sell-side herding ($SLSV < 0$). $LSV_{j,t}$ and $|SLSV_{j,t}|$ are expressed in percentage terms. The number of active investors corresponds to $I_{j,t}$.

These stock months correspond to situations where a large majority of investors take the same position on a given stock. Across the reported cases, the dominant side represents between 55.9% and 100.0% of active investors (mean: 81.5%). In many instances, this corresponds to more than four out of five investors trading in the same direction.²⁶

In some cases, coordination is particularly strong. For example, in December, all active investors buy Caterpillar, while more than 90% buy Newmont Goldcorp and Sipef. These results reflect substantial clustering in trading decisions rather than marginal imbalances.

8.H. Appendix H. Figures

8.H.1. IHM by month Boxplot q05 - q95



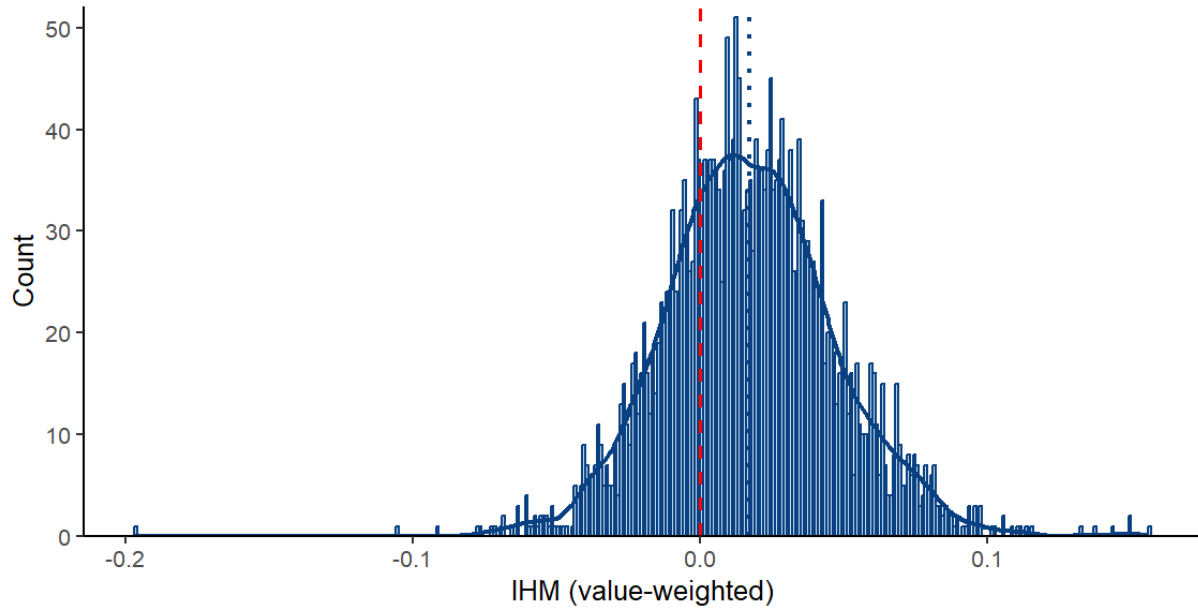
Herding intensity is positive in both months and slightly higher in January than in December. This indicates that coordinated trading behavior strengthens over time, consistent with the increase in stock-level herding ($LSV_{j,t}$) between the two months.

²⁶ The dominant-side share is computed as $B_{j,t} / I_{j,t}$ for buy-side herding and $S_{j,t} / I_{j,t}$ for sell-side herding, where $S_{j,t} = I_{j,t} - B_{j,t}$. Values are based on the stock-month observations reported in Table G4.

8.H.2. IHM Distribution (Histogram + Density): Full range

Figure H2. IHM Distribution (Histogram + Density): Full range

Mean(IHM_vw) = 0.0171

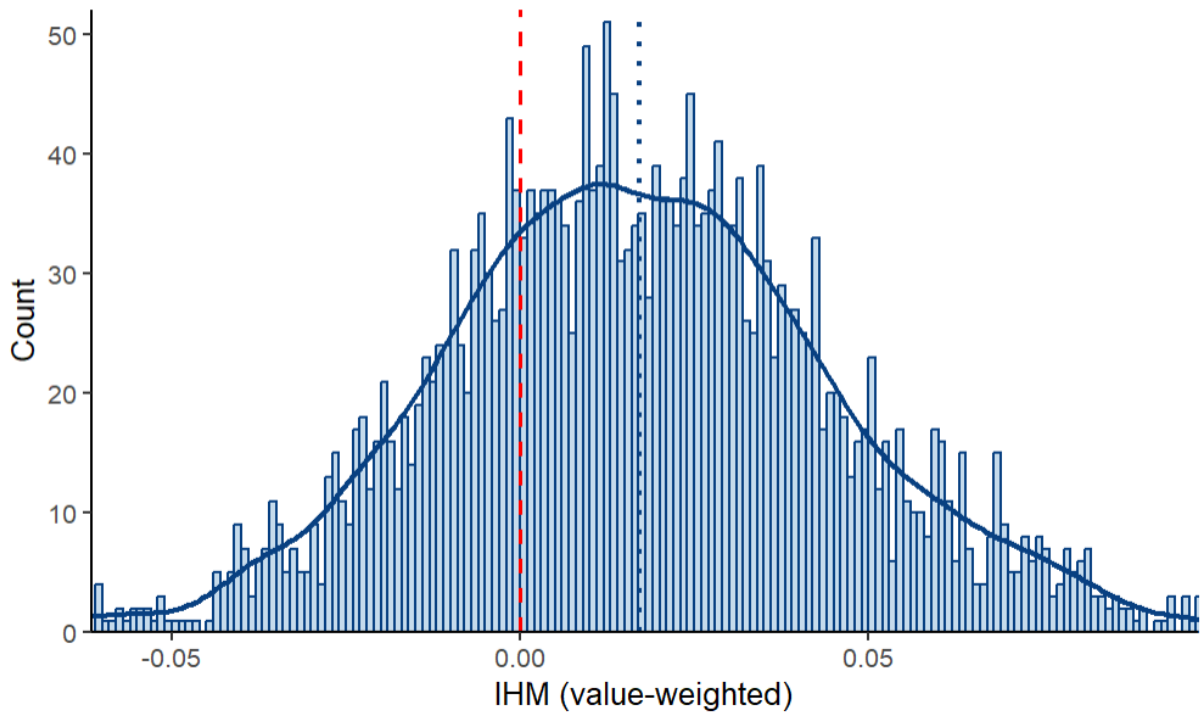


The distribution of $IHM_{i,t}^{vw}$ is centered slightly above zero, with a majority of positive values. This indicates that herding is prevalent among investors, consistent with the rejection of the null hypothesis of no herding.

8.H.3. IHM Distribution (Histogram + Density): q01 - q99

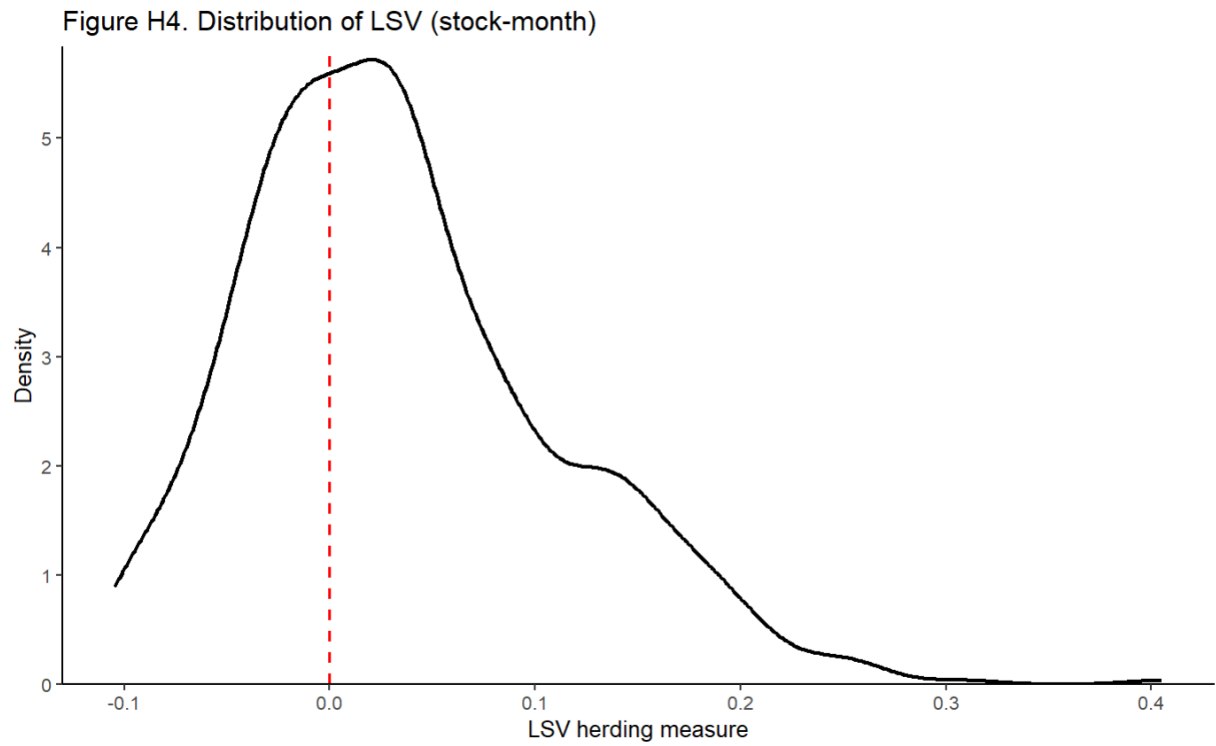
Figure H3. IHM Distribution (Histogram + Density): q01 - q99

Mean(IHM_vw) = 0.0171



To better visualize the central mass of the distribution, *Figure H3* presents a zoomed (q01-q99) version excluding extreme values as the distribution are centered slightly above zero. This indicates that the prevalence of herding is not driven by a small number of extreme observations, but is a widespread feature of investor behavior.

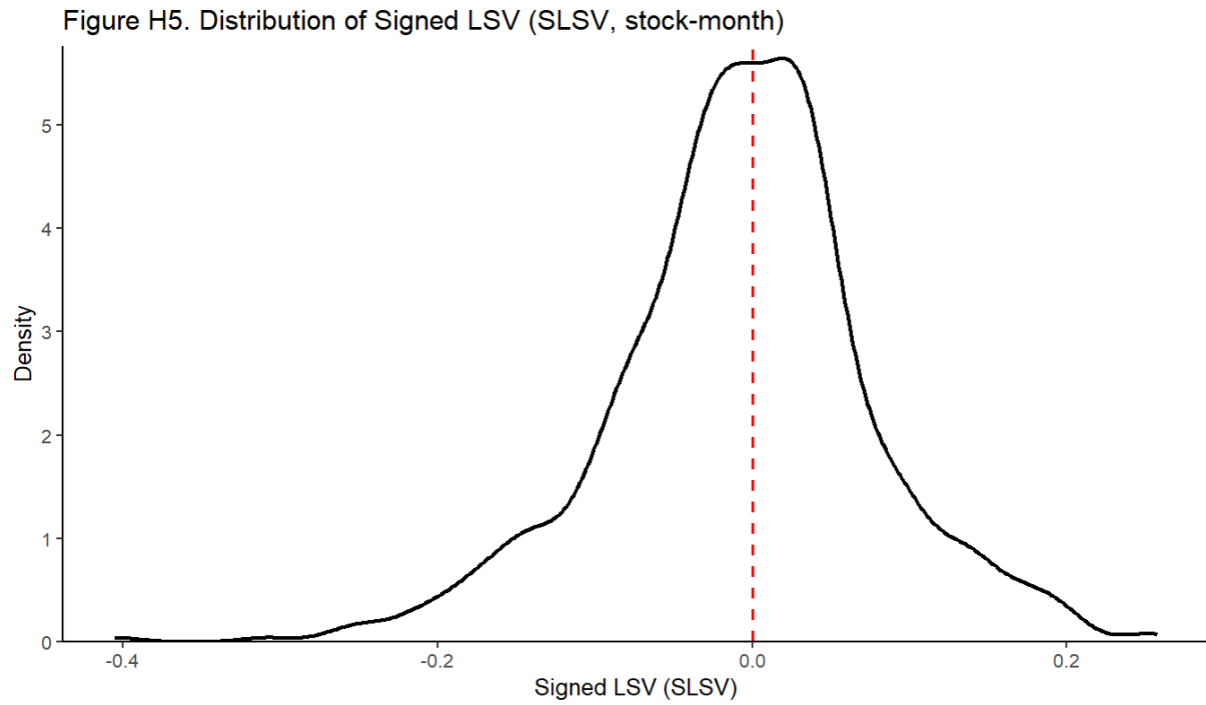
8.H.4. Distribution of LSV (Stock-month)



The distribution of $LSV_{j,t}$ is centered around moderate positive values, indicating the presence of coordinated trading across stocks. The substantial dispersion of the distribution highlights that herding intensity varies considerably across stock-month observations. Values close to zero reflect limited coordination, while higher values indicate stronger synchronization of trading.

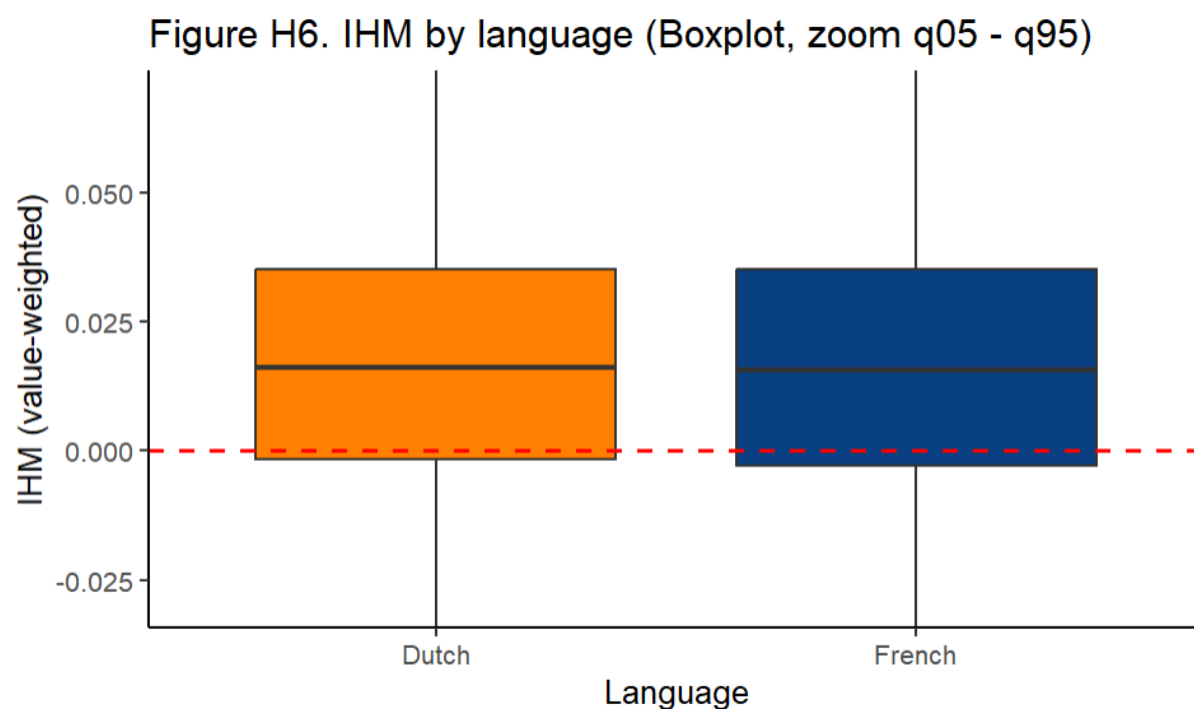
Occasionally, slightly negative values may arise when the observed dispersion is lower than what would be expected under random trading ($|p_{j,t} - p_t| < AF_{j,t}$). These cases reflect an absence of coordination rather than meaningful contrarian behavior.

8.H.5. Distribution of Signed LSV (SLSV, stock-month)



The distribution of $SLSV_{j,t}$ is centered around zero, indicating that herding occurs on both the buy and sell sides. The relatively symmetric distribution suggests that coordinated buying and selling are broadly balanced, with no strong dominance of one direction. This is consistent with the descriptive statistics, which show a slight dominance of sell-side herding.

8.H.6. IHM by language boxplot q05 - q95

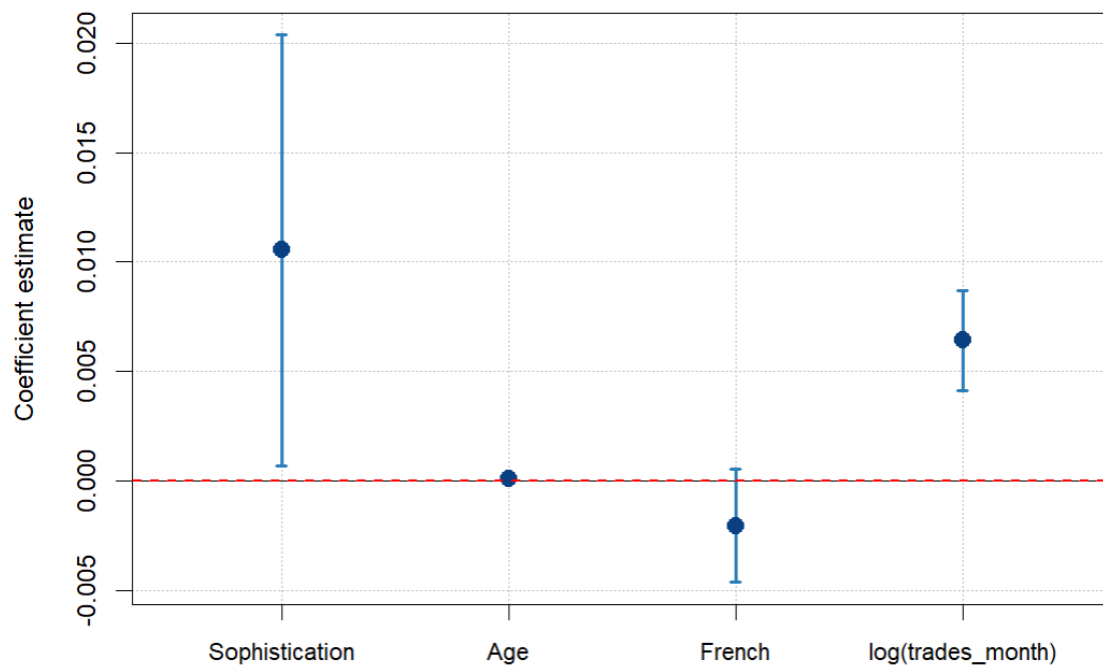


No clear difference in herding intensity is observed between French-speaking and Dutch-speaking investors. The strong overlap in distributions supports the absence of a meaningful relationship between language and herding, consistent with the regression results.

8.H.7. Baseline determinants of herding ($IHM_{i,t}^{vw}$)

Figure H7. Baseline determinants of herding

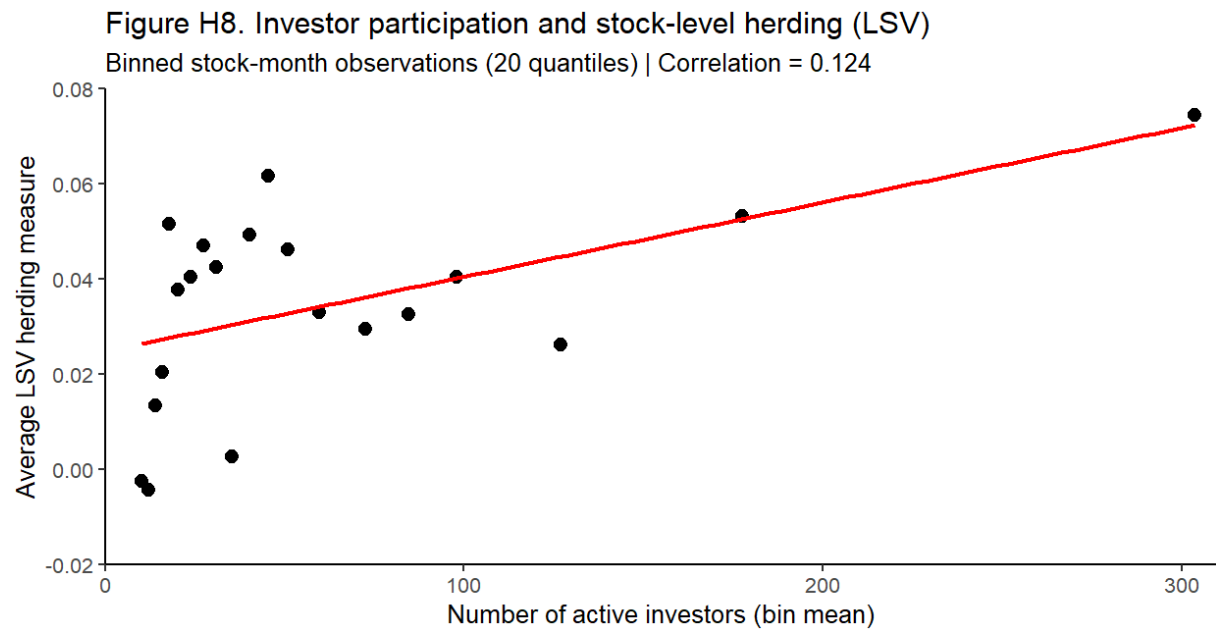
Point estimates with 95% confidence intervals



Notes: Points denote coefficient estimates and vertical bars represent 95% confidence intervals.

The figure summarizes the baseline determinants of herding. Trading activity is the strongest predictor, indicating that more active investors herd more. Financial sophistication is also positively related to herding, while age has a small effect and language shows no meaningful relationship. Confidence intervals indicate which effects are statistically significant.

8.H.8. Investor participation and stock-level herding (LSV)

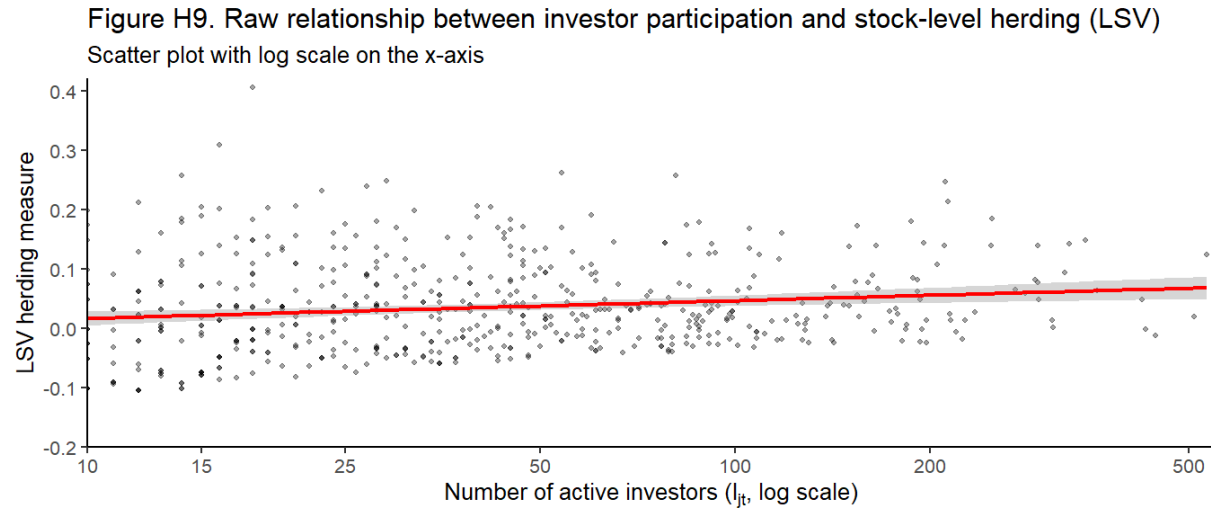


Notes: Each point represents the average herding level within a bin of stock-month observations sorted by investor participation. The red line shows the fitted linear relationship between binned participation and herding (OLS regression).

The figure shows a positive relationship between investor participation and stock-level herding. Herding intensity increases with the number of active investors, suggesting that coordination is stronger in more actively traded stock months.

8.H.9. Investor participation and stock-level herding (LSV): Raw

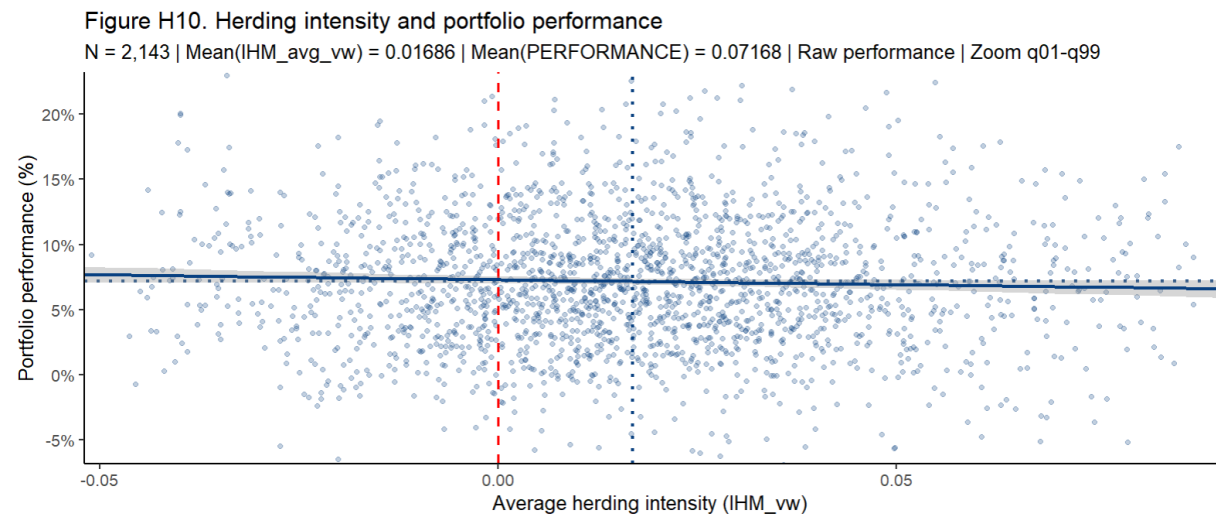
To assess whether this pattern is driven by aggregation, *Figure H9*. presents the raw relationship at the stock-month level.



Notes: The red line shows the fitted linear relationship between investor participation and stock-level herding. The shaded area represents the 95% confidence interval around this estimate. The x-axis is displayed on a logarithmic scale. Because investor participation is highly skewed, the number of investors is expressed in logarithmic form to improve readability.

The raw scatter plot confirms the positive relationship between investor participation and stock-level herding. Although individual observations are widely dispersed, the overall trend remains upward sloping, consistent with the binned representation in *Figure H8*.

8.H.10. Herding intensity and portfolio performance q01 - q99



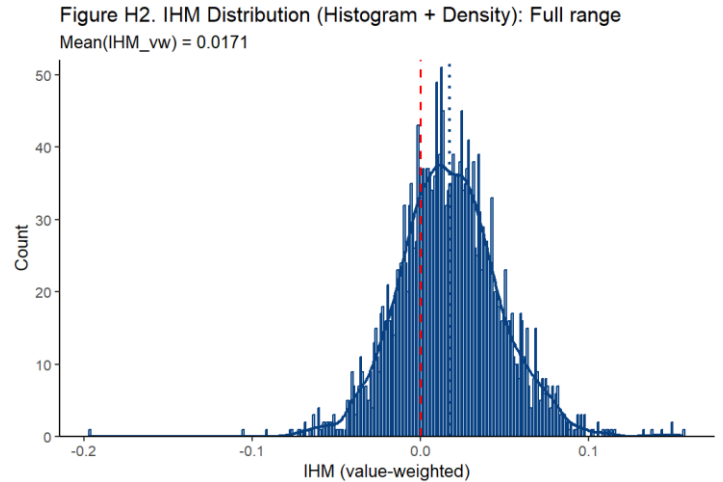
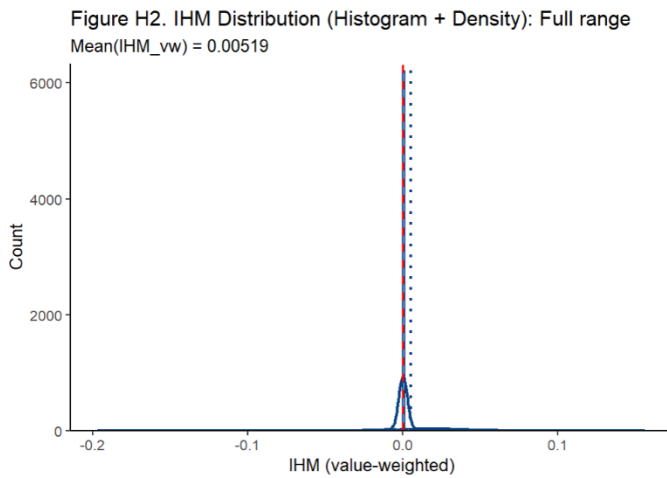
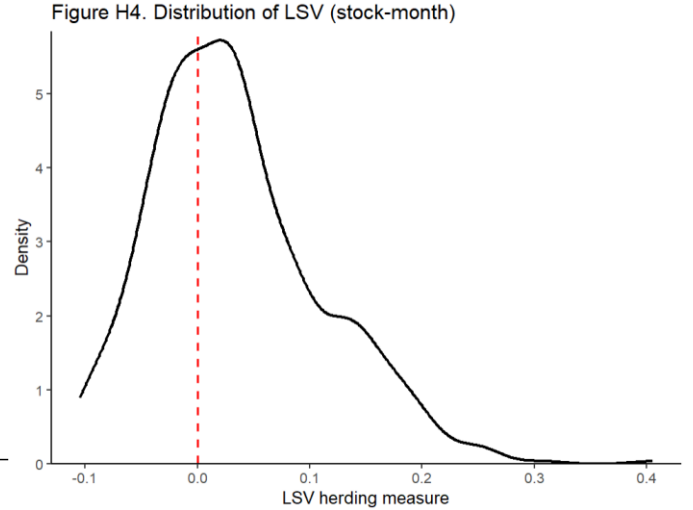
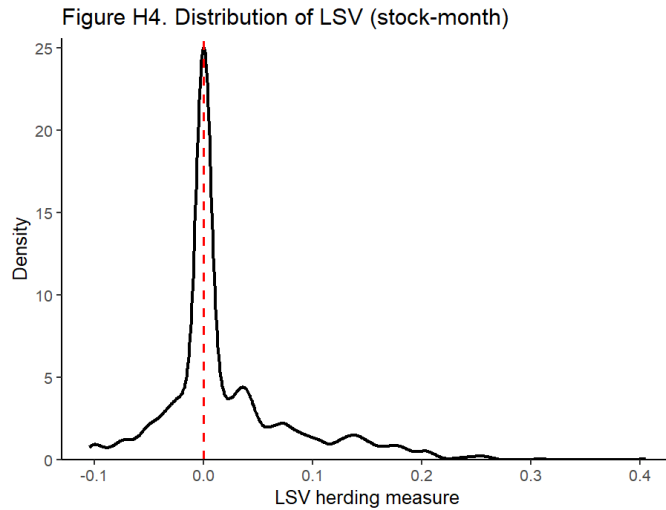
Notes: The fitted line shows the linear relationship between herding intensity and portfolio performance. The shaded area represents the 95% confidence interval around this estimate.

The dotted blue lines indicate sample means.

The figure shows a negative relationship between herding intensity and raw portfolio performance. Investors who follow the crowd tend to achieve lower returns, consistent with the regression results. The pattern remains stable when focusing on the central part of the distribution.

8.H.11. Comparison between full (Nov - Jan) & restricted sample (Dec - Jan)

Those figures compare results from the full three-month sample (on the left) with those from the restricted sample excluding November (on the right).



November clearly compresses both $LSV_{j,t}$ and $IHM_{i,t}^{vw}$ distributions, reflecting the portfolio formation phase during which most investors are net buyers. Excluding this period restores meaningful cross-sectional variation, confirming that November is not informative for measuring herding. The mass of observations around zero is driven by November, where most investors are net buyers.

8.I. Appendix I. Tables

8.I.1. Table II. Winsorization robustness check (performance regression)

Specification (raw performance)	$\beta_{IHM_{i,t}}^{vw}$	Robust SE	R ²	Adjusted R ²	N
Baseline model (no engagement)	-0.1476***	0.0437	0.0898	0.0881	2,143
Winsorized baseline model (1-99%)	-0.1395***	0.0420	0.0912	0.0895	2,143
Final model (with engagement)	-0.2166***	0.0422	0.1502	0.1482	2,143
Winsorized final model (1-99%)	-0.2111***	0.0405	0.1553	0.1533	2,143

Notes: Investor-level regressions of portfolio performance on average herding. Baseline models exclude investor engagement ($LOGINS_i$), while final models include it. Heteroskedasticity-robust (White) standard errors are reported. Performance is winsorized at the 1st and 99th percentiles in the winsorized specifications. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table II. reports the robustness of the performance regression to winsorization of the dependent variable across both baseline and final specifications. The coefficient on herding remains negative and highly significant in all cases, with only a slight reduction in magnitude after winsorization (from -0.2166 to -0.2111). This confirms that the negative relationship between herding and performance is not driven by extreme observations.

Robust (White) standard errors correct for heteroskedasticity, i.e. situations where the variance of regression errors differs across observations. This adjustment does not affect coefficient estimates but ensures that standard errors and statistical significance remain valid.

8.I.2. Table I2. Stock-level herding statistics

Month	Stock-months observed	Stock-months measurable	Mean $LSV_{j,t}$	Buy $SLSV_{j,t}$	Sell $SLSV_{j,t}$
November	337	0	0.0000	-	-
December	334	313	0.0299	47.3%	52.7%
January	337	262	0.0405	46.9%	53.1%
Dec-Jan	671	575 (85.7%)			

Notes: Buy (Sell) $SLSV$ corresponds to the share of stock-month observations exhibiting positive (negative) signed herding. Percentages are computed over the measurable stock-months ($I_{j,t} \geq 10$).

Table I2. reports stock-level herding statistics by month. As expected, no measurable $LSV_{j,t}$ is observed in November, reflecting the portfolio formation phase during which most investors are net buyers, leading to mechanically zero dispersion.

In contrast, herding becomes measurable in December and January, with 85.7% of stock-months satisfying the minimum activity threshold. Mean $LSV_{j,t}$ increases from 0.0299 in December to 0.0405 in January, indicating stronger coordination over time.

Sell-side herding slightly dominates buy-side herding in both months, although the difference remains modest, suggesting broadly balanced directional behavior.

8.I.3. Table I3. LSV stock-month diagnostics

Statistic	Value
Stock-month observations	671
Measurable	575 (85.7%)
Excluded (<10 investors)	14.3%
Mean $LSV_{j,t}$ (Dec)	0.0299
SD $LSV_{j,t}$ (Dec)	0.0707
Mean $LSV_{j,t}$ (Jan)	0.0405
SD $LSV_{j,t}$ (Jan)	0.0837

Note: $LSV_{j,t}$ is computed only for stock-month observations with at least ten active investors ($I_{j,t} \geq 10$). When this condition is not satisfied, the measure is set to N/A (not available) rather than zero, to avoid introducing bias, and these observations are excluded from the analysis.

Table I3. summarizes the coverage and variability of the stock-level herding measure. Out of 671 stock-month observations, 575 (85.7%) are measurable, confirming that LSV is defined for the large majority of cases.

Herding intensity is moderate but non-negligible, with mean LSV equal to 0.0299 in December and 0.0405 in January. The increase between the two months suggests stronger coordination over time, while the relatively large standard deviations (0.0707 and 0.0837) indicate substantial cross-sectional heterogeneity in herding across stocks.

8.I.4. Table I4. Value-weighted vs equal-weighted IHM ($IHM_{i,t}^{vw}$ vs $IHM_{i,t}^{ew}$)

Variable	vw coefficient ($IHM_{i,t}^{vw}$)	ew coefficient ($IHM_{i,t}^{ew}$)
$Sophistication_i$	0.0105**	0.0137***
Age_i	$9.24 \times 10^{-5}***$	$6.73 \times 10^{-5}***$
$French_i$	-0.0021	-0.0027**
$\log(Trades_{i,t})$	0.0064***	0.0063***
Within R ²	0.0229	0.0276

Notes: Standard errors clustered at the investor level (*USER_ID*).

*, **, *** denote significance at the 10%, 5%, and 1% levels.

Table I4. compares the value-weighted and equal-weighted specifications of the IHM. The results are qualitatively similar across both measures: financial sophistication and trading activity are positively related to herding, while age remains positive but modest in magnitude.

Differences appear mainly in statistical strength and fit, with the equal-weighted specification yielding a slightly higher Within R² (0.0276 vs 0.0229). This reflects a better fit to directional trading patterns rather than economic importance.

Since $IHM_{i,t}^{vw}$ accounts for the monetary size of trades, it is retained as the preferred baseline measure.

8.I.5. Table I5. Level vs Log specification

Specification	R ²	Within R ²	β ($\log(\text{Trades}_{i,t})$)
vw - Level	0.0188	0.0163	0.0001
vw - Log	0.0253	0.0229	0.0064
ew - Level	0.0233	0.0197	0.0001
ew - Log	0.0311	0.0276	0.0063

Notes: Standard errors clustered at the investor level (*USER_ID*).

*, **, *** denote significance at the 10%, 5%, and 1% levels.

Table I5. compares the level and logarithmic specifications of trading activity. The log specification leads to a clear improvement in model fit in both cases: in the value-weighted model ($IHM_{i,t}^{vw}$), R² increases from 0.0188 to 0.0253 (Within R² from 0.0163 to 0.0229), and in the equal-weighted model ($IHM_{i,t}^{ew}$) from 0.0233 to 0.0311 (Within R² from 0.0197 to 0.0276), corresponding to gains of around 30 to 35%.

At the same time, coefficient signs and significance remain unchanged, while the coefficient on trading activity becomes economically more interpretable under the log form (≈ 0.0064). This supports the use of the logarithmic specification, which captures diminishing marginal effects of trading activity and provides a better empirical fit.

8.J. Appendix J. R Code

The R script is provided in a separate document.

9. Bibliography

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review*, 87(5), 847-862. <https://www.jstor.org/stable/2951328>
- Andrikopoulos, P., Kallinterakis, V., Ferreira, M. P. L., & Verousis, T. (2017). *Intraday herding on a cross-border exchange*. *International Review of Financial Analysis*, 53, 25-36. <https://doi.org/10.1016/j.irfa.2017.08.010>
- Avramov, D., Chordia, T., & Goyal, A. (2006). Liquidity and autocorrelations in individual stock returns. *The Journal of Finance*, 61(5), 2365-2394. <https://doi.org/10.1111/j.1540-6261.2006.01060.x>
- Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797-817. <https://doi.org/10.2307/2118364>
- Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261-292. <https://doi.org/10.1162/003355301556400>
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21(2), 785-818. <https://doi.org/10.1093/rfs/hhm079>
- Bellofatto, A. (2017). *How does language impact foreign investing in a multilingual country?* [Working Paper]. Université Catholique de Louvain, Louvain School of Management. <https://www.researchgate.net/publication/318395117>
- Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1), 23-36. <https://doi.org/10.2307/1909829> (Original work published 1738 in Latin: <https://archive.org/details/SpecimenTheoriaeNovaeDeMensuraSortis/page/n11/mode/2up>)
- Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets. *IMF Staff Papers*, 47(3), 279-310. <https://www.imf.org/external/pubs/ft/staffp/2001/01/pdf/bikhchan.pdf>
- Blasco, N., Corredor, P., & Ferreruela, S. (2009). The implications of herding on volatility: The case of the Spanish stock market. *Working Paper*, Universidad Pública de Navarra. <https://archivo.alde.es/encuentros.alde.es/anteriores/xiieea/trabajos/pdf/095.pdf>
- Boosi, R. (2025). *A critical analysis of eToro copy trading risks and GameStop's short squeeze: Anomalous market conditions and the perils of speculative investment*. Preprint. https://www.researchgate.net/profile/Roe-Boosi/publication/392082820_Roe_Boosi_A_Critical_Analysis_of_eToro_Copy_Trading_Risks_and_GameStop%27s_Short_Squeeze-Anomalous-Market_Conditions_and_the_Perils_of_Speculative_Investment.pdf
- Broihanne, M.-H., & Capelle-Blancard, G. (2018). Richard Thaler ou comment la finance est devenue comportementale. *Revue d'économie politique*, 128(4), 549-574.

<https://doi.org/10.3917/redp.284.0549>

Brown, N. C., Wei, K. C. J., & Wermers, R. (2007). *Analyst recommendations, mutual fund herding, and overreaction in stock prices* (CFR Working Paper No. 07-08). Centre for Financial Research, University of Cologne.
<https://www.econstor.eu/bitstream/10419/57711/1/703027662.pdf>

Brunnermeier, M. K. (2001). *Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding*. Oxford University Press.
<https://doi.org/10.1093/0198296983.001.0001>

Brunnermeier, M. K. (2017). *Bubbles*. Princeton University Working Paper.
https://www.princeton.edu/~markus/research/papers/bubbles_survey.pdf

Cauchies, C. (2014). *Herding behavior in the limit order book* [Master's thesis, Université Catholique de Louvain, Louvain School of Management]. UCLouvain Repository.
<https://thesis.dial.uclouvain.be/entities/masterthesis/df56898b-ae0d-4f48-9c97-b00b04570496>

De Schrijver, J. (2022). *Analyse de l'effet de disposition parmi les participants du Rallye Boursier de l'Echo* [Master's thesis, Université catholique de Louvain, Louvain School of Management]. UCLouvain Repository. <https://hdl.handle.net/2078.1/thesis:35055>

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, 40(3-5), 603-615. [https://doi.org/10.1016/0014-2921\(95\)00073-9](https://doi.org/10.1016/0014-2921(95)00073-9)

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>

Farhath, A. (2025). Take-Two delays GTA 6 release to November 2026; shares fall. *Seeking Alpha*. <https://seekingalpha.com/news/4517840-take-two-delays-gta-6-release-to-november-2026-shares-fall>

Freibauer, J., Grawert, S., & Rieger, M. O. (2024). The effects of trading apps on investment behavior over time. *The European Journal of Finance*. Advance online publication. <https://doi.org/10.1080/1351847X.2024.2401604>

Frey, S., Herbst, P., & Walter, A. (2014). Measuring mutual fund herding: A structural approach. *Journal of International Financial Markets, Institutions & Money*, 32, 219-239. <https://doi.org/10.1016/j.intfin.2014.05.006>

Grant, K. (2021). GameStop: What is it and why is it trending? *BBC Newsbeat*. <https://www.bbc.com/news/newsbeat-55841719>

Harring, A. (2023). How Tesla rose to retail investor stardom: 'It's always in people's minds.' *CNBC*. <https://www.cnbc.com/2023/12/20/how-tesla-rose-to-retail-investor-stardom-its-always-in-peoples-minds.html>

Hasanah, E. N., Angelica, A. E., Koesrindartoto, D. P., & Wiryono, S. K. (2025). Who deserves to be the finfluencer? *Journal of Open Innovation: Technology, Market, and Complexity*, 11, 100553. <https://doi.org/10.1016/j.joitmc.2025.100553>

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*.

Sage Publications.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage Publications.

Investing.com. (2025). Record retail buying of Nvidia shares following AI concerns. <https://www.investing.com/news/stock-market-news/record-retail-buying-of-nvidia-shares-following-ai-concerns-93CH-3833733>

Jurkatis, S. (2022). Why you should not use the LSV herding measure (Bank of England Staff Working Paper No. 959). *Bank of England*. <https://www.bankofengland.co.uk/working-paper/2022/why-you-should-not-use-the-lsv-herding-measure>

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus & Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-292. <https://doi.org/10.2307/1914185>

Kahneman, D., & Tversky, A. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323. <https://doi.org/10.1007/BF00122574>

Korniotis, G. M., & Kumar, A. (2011). Do older investors make better investment decisions? *The Review of Economics and Statistics*, 93(1), 244-265. https://doi.org/10.1162/REST_a_00053

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1), 23-43. [https://doi.org/10.1016/0304-405X\(92\)90023-Q](https://doi.org/10.1016/0304-405X(92)90023-Q)

L'Echo. (2025). Rallye Boursier 2024-2025. <https://www.lecho.be/rallyeboursier.html>

Lobão, J., & Maio, J. (2019). Herding around the world: Do cultural differences influence investors' behavior? *Research in International Business and Finance*, 49, 319-333. <https://doi.org/10.1016/j.ribaf.2019.03.007>

MarketWatch. (2025). GameStop Corp. Cl A (GME) company profile. <https://www.marketwatch.com/investing/stock/gme/company-profile>

Merli, M., & Roger, T. (2013). What drives the herding behavior of individual investors? *Finance*, 34(3), 67-104. <https://doi.org/10.3917/fina.343.0067>

Nofsinger, J. R. (2022). *The psychology of investing* (7th ed.). Routledge. <https://doi.org/10.4324/9781003159704>

Olicard, F. (2021). *Votre attention est votre super-pouvoir*. First Éditions.

Pascal, B. (1671). *Pensées* (3rd ed.). Guillaume Desprez. https://www.samizdat.qc.ca/arts/lit/Pascal/Pensees_1671_ancien.pdf

Pomparelli, T. (2024). The GameStop short squeeze as a case study in business law education. *Education Quarterly Reviews*, 7(2), 199-206. <https://doi.org/10.31014/aior.1993.07.02.589>

Renard, H. (2025). *Les participants au Rallye Boursier de L'Echo adoptent-ils une stratégie contrarian ou momentum ?* [Master's thesis, Université catholique de Louvain, Louvain School of Management].

Seitz, B., Rathgeber, A., Lameli, A., & Gaugler, T. (2017). *Sharing more than dialects: Herding behavior in bank risk-taking explained by cultural vicinity*. SSRN. <https://doi.org/10.2139/ssrn.2909878>

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4), 967-1004. <https://doi.org/10.1257/aer.107.4.967>

Singh, V. (2013). Did institutions herd during the Internet bubble? *Review of Quantitative Finance and Accounting*, 41(3), 513-534. <https://doi.org/10.1007/s11156-012-0320-1>

Sismanis, N. (2025). Take-Two Interactive: GTA VI may crush it, but taking profits is wise. *Seeking Alpha*. <https://www.seekingalpha.com/article/4802455-take-two-interactive-gta-vi-may-crush-it-but-taking-profits-is-wise>

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214. <https://doi.org/10.1287/mksc.4.3.199>

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3), 183-206. [https://doi.org/10.1002/\(SICI\)1099-0771\(199909\)12:3%3C183::AID-BDM318%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0771(199909)12:3%3C183::AID-BDM318%3E3.0.CO;2-F)

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458. <https://doi.org/10.1126/science.7455683>

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press. <https://uvammm.github.io/docs/theoryofgames.pdf>

Wang, G., & Wang, Y. (2018). Herding, social networks, and volatility. *Economic Modelling*, 68, 74-81. <https://doi.org/10.1016/j.econmod.2017.04.018>

Wheat, C., & Eckerd, G. (2025). *Crypto investor waves since 2017: What retail investor behavior reveals about digital asset adoption*. JPMorgan Chase Institute. <https://www.jpmorganchase.com/institute/all-topics/financial-health-wealth-creation/crypto-investor-waves-since-2017-what-retail-investor-behavior-reveals-about-digital-asset-adoption>

Zheng, Z., Tang, K., Liu, Y., & Guo, J. M. (2021). Gender and herding. *Journal of Empirical Finance*, 64, 1-21. <https://doi.org/10.1016/j.jempfin.2021.10.005>

Abstract:

In 2021, the GameStop episode demonstrated the growing influence of individual investors on financial markets. Retail investors, coordinated through online communities, managed to trigger a historic short squeeze from institutional investors, driving stock prices upward. Such events show that understanding the behavior of individual investors is not only relevant for academic research but also for market participants, especially those seeking to anticipate or manage volatility. Moreover, the rise of social media platforms and 'influencers' amplifies imitation patterns, commonly referred to as herding.

Despite this evidence, traditional finance models continue to rely on the assumption of investor rationality. In the 1970s, Eugene Fama's EMH argued that asset prices reflect all available information, implying that persistent deviations from fundamental values should not occur and that assets are fairly priced. However, events like the GameStop one show that markets become inefficient when collective biases dominate, emphasizing the importance of behavioral finance. Research on herding behavior among both institutional and retail investors has grown considerably in recent years. Although rational investors should not neglect their own information, a significant share of them actually does. They tend not only to imitate others' strategies, but also to replicate their overall behavior.

This thesis is based on behavioral finance literature and established herding measures, notably the Lakonishok, Shleifer and Vishny measure (LSV) and the Individual Herding Measure (IHM). The study by Merli and Roger (2013) employing the IHM also significantly shapes this thesis. The originality of this paper lies in linking classic herding metrics, such as LSV, with detailed investor data from the *Rallye Boursier* and contemporary market dynamics. Herding patterns are analyzed using socio-demographic, cultural, portfolio-related and stock-level variables. In addition, this empirical study combines behavioral theories and modern investor characteristics to contribute to the literature on retail investor herding.