Category Learning, Trade Volume, and the Advent of AI

Ramiro H. Gálvez[a], Paula Margaretic[b], Rafael Paganini[c], Pablo A. Macri[d],

September 2024

Abstract

This paper presents a noisy rational expectations model in which attention-constrained investors allocate their attention between asset-specific fundamentals and categorical factors influencing groups of assets. The model predicts that more attention is directed toward larger categories, which leads to increased trade volumes in assets belonging to those categories. We empirically validate these predictions by analyzing data from the social media platform X (formerly Twitter), focusing on how users mention and co-mention assets over time. We construct a dynamic network of co-mentions and use community detection algorithms to identify perceived similarities between assets. Additionally, we explore the impact of artificial intelligence (AI) on investors' behavior, showing that AI enhances categorical learning while also increasing selectivity in community formation. Our results demonstrate that AI influences trade volume and volatility, with significant implications for understanding the role of AI in reshaping market dynamics and investor behavior.

Keywords: attention allocation, categorical learning, trade volume, artificial intelligence

[a] Universidad Torcuato Di Tella, Email: rgalvez@utdt.edu

[b] Universidad de Chile, Facultad de Economía y Negocios. Email: pmargareti@fen.uchile.cl

[c] Universidad de San Andrés, Email: rpaganinipesce@udesa.edu.ar

[d] Universidad de San Andrés, Email: pmacri@udesa.edu.ar

1 Introduction

Big Data (BD) and Artificial Intelligence (AI) have transformed every aspect of the financial sector, including trading, brokerage services, and banking (Koijen and Yogo, 2022).[1] BD and AI minimize the role of human judgment in generating usable knowledge (Abis and Veldkamp, 2024). In most sectors, technological progress boosts efficiency. However, in finance, the proliferation of data, advancements in data processing, and the rise of data-driven trading strategies have been linked to market volatility, illiquidity (Arifovic et al., 2022), and displacement of resources towards rent-seeking activities, (Glode and Ordoñez, 2024). As discussed by Farboodi and Veldkamp (2020), one reason financial technology is suspect is that its rise has been accompanied by a shift in the nature of financial analysis and trading. Rather than using these vast datasets for more detailed analysis, many traders focus on statistical arbitrage, demand data (Farboodi and Veldkamp, 2020), and style investments (Barberis and Shleifer, 2003), developing algorithms that exploit patterns in order flows and capitalize on "dumb money" from others' trades.

Key questions arise: How do investors choose between asset-specific and market fundamental information? Why are their incentives to process each type of data changing? How do investors incorporate other traders' views? Answering these questions requires a model (Farboodi and Veldkamp, 2020) that connects it with the data. A large body of research highlights the role of attention in how agents and prices process information. Attention, a scarce resource (Bose et al., 2022), significantly influences investment decisions and financial markets (Barber and Odean, 2008, Hillert et al., 2014, Mohrschladt and Schneider, 2021). This is especially relevant in the information age, where an abundance of information leads to a scarcity of attention (Da et al., 2011).

We begin by developing a noisy rational expectations model where attention-constrained investors must allocate their attention between asset-specific and categorical fundamentals affecting a subset of assets. Next, we propose empirical measures for these forms of attention using web traffic data. Specifically, we examine the social network X to extract users' collective perception of the apparent similarity among stocks. The construction of our measures relies on two key intuitions. First, when users mention one asset or two or more assets together, they are directing attention to these assets. Second, co-mentioning suggests that users likely perceive these assets as sharing common characteristics or categorical factors. Utilizing natural language processing techniques (Gentzkow et al., 2019) and sentiment analysis, we extract valuable insights from these unstructured, large datasets, enabling us to implement our proposed measures. We further validate our measures by testing the primary theoretical implications of our model using real-world data. Notably, our measures yield results consistent with actual economic outcomes. Finally, we examine the effect of artificial intelligence (AI) on category-

[1]The digital age has produced massive amounts of data, which continue to grow exponentially. The International Data Corporation estimates that the world generates more data in two days than humanity produced from the dawn of time to 2003 (Goldstein et al., 2021).

learning behavior, trade volume, and trade volume volatility. In our model, we treat the rise of AI as an expansion in investors' attention capacity, building on existing literature that explores the influence of advancements in information technology (Begenau et al., 2018, Farboodi and Veldkamp, 2020).

We combine a big dataset of social media posts from the platform X (formerly Twitter) to empirically measure attention towards categorical and asset-specific factors. Our dataset consists of over 22 million X posts between January 2015 and December 2022. These posts mention cashtags, which represent specific stocks or assets. Each post contains at least one cashtag associated with stocks in the S&P 500 index at a given date-hour.[2] From this dataset, we construct time-varying *networks of cashtags* that capture how different assets are co-mentioned on X at specific points in time (an hour and a date). In these networks, nodes represent assets, and edges represent co-mentions of assets within a post. The weight of each edge reflects the number of times two assets are mentioned together, suggesting their perceived similarity.

To identify communities within the cashtag network at each time period, we apply a community detection algorithm (Blondel et al., 2008). These communities serve as empirical counterparts to the categories in our theoretical model. By tracking the number of mentions and co-mentions of assets within these communities over time, we derive proxies for category-level and asset-specific attention. Additionally, our analysis includes AI-related measures based on search trends and social media data to gauge AI's rising importance. Lastly, we utilize intraday volume data from the EOD Historical Data API to study trade volume for the S&P 500 stocks. This high-frequency data enables us to examine the relationship between our attention measures and market activities such as trade volume and volatility.

Our theoretical framework shows that attention-constrained investors allocate more attention to categorical factors relative to asset-specific fundamentals, leading to category learning behavior. Intuitively, since categorical signals contain information relevant to a group of assets, whereas asset-specific signals are limited to individual assets, investors prefer to prioritize the former when they are attention-constrained. Furthermore, category learning increases with the size of the categories. The reason for this result is that the informational benefit of category-level signals is higher for more central categories —those containing a larger number of assets— thus leading investors to concentrate their attention on these signals. We investigate the implications of these findings for trade, showing that investors trade more in assets about which they hold more precise beliefs, and they form these precise beliefs by focusing more attention on those assets. Finally, we demonstrate that trade volume is higher for larger categories where category-learning behavior is more prevalent. We validate our measures by testing the primary theoretical implications of our model using our real-world data. Notably, our measures yield results consistent with actual economic outcomes.

[2]Importantly, focusing on posts that mention S&P 500 stocks does not restrict our analysis solely to these stocks, as posts may also include cashtags from assets not directly related to the S&P 500, such as cryptocurrencies.

Finally, using our AI proxies, we examine the relationship between AI, categorical attention and various market outcomes, focusing on trade volume and its volatility. The analysis provides important insights into the role of AI in reshaping investor behavior and market dynamics. We find that AI positively impacts our proxy for categorical attention. Interestingly, we observe a regime shift in categorical attention levels starting in 2021, which corresponds to the surge in AI technologies. We characterize this regime shift with a higher level and standard deviation of our proxy for categorical attention.

The way we interpret this sort of regime shif in our measure is twofold. On the one hand, an increase in AI (which, in the model, corresponds to a relaxation of the attention constraint) enables investors to allocate greater attention to both categorical (and asset-specific) factors, thereby forming more accurate beliefs regarding asset payoffs. On the other hand, rising AI also allows investors to become more selective (and occasionally erratic) in forming communities of perceived similar assets (i.e., those frequently mentioned together). This increased selectivity likely accounts for the observed rise in the standard deviation of our measure. Finally, we find a negative correlation between our AI measures and overall trade volume but a positive relation with trade volume volatility.

To sum up, our paper not only introduces novel measures of attention allocation in financial markets but also offers an empirical investigation into the complex interplay between investor behavior, and market outcomes. As an application, we consider the impact of AI investor behavior and trade volume. Our findings suggest that AI, while enhancing information processing capabilities, also alters traditional investment dynamics, with considerable implications for trade volume and market volatility. This research adds to the broader understanding of how AI and big data are reshaping the financial landscape, offering both theoretical and practical insights into the future of trading and investment strategies.

The structure of the paper follows. Section 2 reviews the literature. Section 3 presents the theoretical model, along with its main theoretical results. Section 4 details the data, while Section 5 presents a validation analysis of our proposed measures. Section 6 examines the relationship between categorical attention and trade volume in the era of IA. Section 7 concludes with some venues of future research. We relegate additional analyses, robustness checks, and the proofs of the theoretical propositions and theorems to the Appendix.

2 Literature review

There is a rich theoretical and empirical literature supporting the importance of attention. While influential models where information acquisition is costly, such as Grossman and Stiglitz (1980), have existed for several years now, the Rational Inattention literature, initiated by Sims (2003), uniquely centers attention as the scarce resource linked to such costs. This perspective has enhanced our understanding of various financial market dynamics (see Maćkowiak et al. (2023) for a review), including home bias (Van Nieuwerburgh and Veldkamp, 2009, Mon-

dria et al., 2010), lack of diversification (Van Nieuwerburgh and Veldkamp, 2010), asset price co-movement (Peng and Xiong, 2006, Mondria, 2010), and financial contagion (Mondria and Quintana-Domeque, 2013).

In particular, attention allocation models that incorporate sector- or market-level fundamental factors, which are relevant to multiple assets simultaneously, have been used to analyze price reactions to earnings announcements of firms with varying exposure to systematic risk factors (Andrei and Hasler, 2015). These models also help explain the relative abundance of data about large, high-growth firms (Farboodi et al., 2022). Unlike models where private signals are either known or associated with public signals equally observed by all participants, some models incorporate private signals into the attention allocation problem. For instance, Kacperczyk et al. (2016) explores the value of informed mutual fund managers, while Peng and Xiong (2006) explains phenomena such as excessive return comovement through category learning, framed as an optimal strategy for attention-constrained investors.

A growing body of empirical evidence suggests that attention influences price movements, returns, risk, and liquidity, enabling profits for some investors (Hirshleifer et al., 2011, Da et al., 2011, Hao and Xiong, 2021, Chakrabarty and Moulton, 2012, Ben-Rephael et al., 2017, Gargano and Rossi, 2018). For instance, Da et al. (2011) construct an attention measure based on Google search frequency, finding it is predictive of short-term price increases and longer-term reversals. This measure likely reflects retail investor behavior. Conversely, Ben-Rephael et al. (2017) use Bloomberg terminal data, capturing institutional investor attention. They find that institutional attention leads retail attention, offering an explanation for price drifts following events like earnings announcements.

Additionally, Gargano and Rossi (2018) use brokerage account data to measure individual investor attentiveness, showing that more attentive investors tend to achieve higher returns on stocks they buy early. Relevant to our model, Da and Warachka (2011) find that both returns and revisions in long-term earnings forecasts predict the disparity between long-term and short-term forecasts. They argue that this reflects how investors rely more on categorical classifications, such as a stock's previous classification as growth or value, while paying less attention to firm-specific heterogeneity—a belief-updating process similar to the one in our model.

We contribute to this literature by proposing an empirical measure of category learning behavior, derived from a noisy rational expectations model in which attention constraints drive this behavior. Unlike Peng and Xiong (2006) and much of the empirical literature on attention, which primarily examines its impact on prices and returns, we investigate both theoretically and empirically category learning behavior and how it affects trade volume. Our measures of category-level and asset-specific attention leverage the big data available from X, using NLP machine learning techniques and sentiment analysis algorithms to extract textual information from posts. To quantify category-level attention, we identify users' co-mentioning patterns on the social network, which allows us to infer their collective perceptions of how similar two or

more assets are. The empirical analysis strongly supports our theoretical findings, indicating that our proposed measures effectively capture category learning behavior and asset-specific attention by producing results consistent with actual economic outcomes.

As mentioned, the paper most closely related to ours is Peng and Xiong (2006). We diverge from Peng and Xiong (2006) in three key ways. First, while we use a noisy rational expectations equilibrium framework with risk-neutral investors to examine the impact of category learning on trade volume, Peng and Xiong (2006) employs a representative agent who optimally makes intertemporal consumption decisions based on her beliefs about fundamental factors. In their model, asset prices are determined by the pricing kernel, linked to the investor's marginal utility of consumption. Their equilibrium involves the investor allocating attention across fundamental factors to minimize total portfolio uncertainty, with asset prices fluctuating as the investor updates her beliefs. Second, in Peng and Xiong (2006), both market-level and category-level factors are present, whereas our model focuses solely on category-level factors, with investors receiving noisy private signals about them. We exclude the market-wide factor for tractability, but this still allows for the emergence of category-related learning. Finally, Peng and Xiong (2006) introduces an overconfidence bias, which they model as the investor's exaggerated belief in her information-processing ability.

This paper also contributes to the literature on the impact of big data and AI on the economy and financial markets (Begenau et al., 2018, Farboodi and Veldkamp, 2020, Arifovic et al., 2022). Part of this literature seeks to understand the market effects of changes in agents' behavior induced by these recent developments. For example, Farboodi and Veldkamp (2020) develop a model where, at least during a transitional phase, increased data processing capabilities shift information acquisition away from fundamentals and toward other investors' demands. Similarly, Arifovic et al. (2022) simulate a market with high-frequency algorithmic traders who use machine learning, documenting negative effects on liquidity.

Another related strand of literature explores the implications of changes in the economy's information structure due to greater access to data and enhanced data processing capabilities. For instance, Abrardi et al. (2020) and Brunnermeier et al. (2020) model insurance markets where the informational advantage lies with insurers rather than policyholders, as insurers' algorithmic assessments of risk may be more accurate than those of policyholders themselves. Begenau et al. (2018) propose a model that rationalizes the idea that big data disproportionately benefits larger firms. These firms generate more data, enabling investors to better forecast their stock returns, which reduces the firms' risk premiums. This creates a feedback loop where the lower cost of capital fuels further growth. We contribute to this literature by exploring the evolution of category learning behavior and its relationship with artificial intelligence, which we model as an increase in the attention endowment.

Finally, our empirical approach makes a valuable contribution to the growing literature on understanding the intelligence embedded in web traffic data (Da and Warachka, 2011, Lee et al., 2015, Ben-Rephael et al., 2017, Fedyk, 2021, Benamar et al., 2021). While existing

studies, such as those using Google Trends, typically measure the nominal level of traffic at the firm level as a proxy for attention, our methodology shifts the analysis to the user level. The higher granularity of our data enables us to identify users' co-mentioning patterns. By being one of the first to analyze traffic patterns with this level of detail and combining it with machine learning techniques to handle large datasets, we contribute to the literature by uncovering patterns that simpler dataset cannot reveal (Goldstein et al., 2021).

3 The model

3.1 Setup

Consider an economy with a continuum of investors $i \in [0, 1]$, uniformly distributed along the unit interval, who make portfolio decisions in a three-period game, with $t \in \{0, 1, 2\}$ denoting the period. Let W_t^i be investor i's wealth in period t. All investors have the same initial wealth W_0. In periods $t = 0$ and $t = 1$, investors decide how to allocate their wealth between a risk-free asset and a set of n risky securities $n \in \{1, \ldots, N\}$. Risky securities belong to one of two categories $m \in \{A, B\}$. Let N_A and N_B represent the number of assets in category A and B, respectively, such that total number of assets in this economy is $N_A + N_B = N$. Asset categories may reflect differences in risk profiles, industry sectors, or other characteristics pertinent to investors' decision-making processes. For simplicity, and without loss of generality, we restrict to two categories. We write $m(n)$ to denote the category m to which asset n belongs, with $m(n) = A$ if $n \in \{1, \ldots, N_A\}$ and $m(n) = B$ if $n \in \{N_A + 1, \ldots, N\}$.

Each unit of risky asset n delivers a final payoff $D_n = v_n + c_m$ in period $t = 2$ to any investor that holds it in her portfolio, where v_n is an asset-specific fundamental factor and c_m is a fundamental factor that is common to all assets in the same category (i.e., assets n such that $m(n) = m$). Investor i has prior beliefs about v_n and c_m at $t = 0$, which are represented by distributions $\tilde{v}_{n0}^i$ and $\tilde{c}_{m0}^i$. The risk-free asset has an associated payoff of 1 for each unit (that is, there is an exogenous risk-free rate of return of 0). Each investor is solely interested in her final wealth in period $t = 2$, and has preferences represented by a CARA utility function with a risk-aversion parameter of 1:

$$u_i(W_2^i) = -e^{-W_2^i}$$

Denote by X_{nt}^i and Y_n^i the quantities of the risky and risk-free assets that investor i chooses to hold at period t. Let P_{nt} be the market price for the risky asset n in period t. The equilibrium price for the risk-free asset is 1. At $t = 2$, investor's wealth must equal her holdings of the risky and risk-free assets as given by

$$W_2^i = Y_1^i + \sum_{n=1}^{N} D_n X_{n1}^i$$

We allow for negative quantities, implying short positions of the asset in question. It is convenient to replace Y_1^i in the previous equation with an expression that reflects the investor's

resource constraint in $t = 1$: $W_1^i = Y_1^i + \sum_{n=1}^N P_{n1} X_{n1}^i$. We then have

$$W_2^i = \left(W_1^i - \sum_{n=1}^N P_{n1} X_{n1}^i \right) + \sum_{n=1}^N D_n X_{n1}^i$$

which in turn implies

$$W_2^i = W_1^i + \sum_{n=1}^N \left(D_n - P_{n1} \right) X_{n1}^i \tag{1}$$

Analogously, wealth in $t = 1$, W_1^i, is given by

$$W_1^i = W_0^i + \sum_{n=1}^N \left(P_{n1} - P_{n0} \right) X_{n0}^i \tag{2}$$

For all risky assets n, net supply for each asset is set at 0 at $t = 0$ and follows $a_n \sim \mathcal{N}\left(0, \frac{1}{q_a}\right)$ at $t = 1$. We assume the net supplies a_n and $a_{n'}$ are independent whenever $n \neq n'$. Investors know the distribution of a_n but they do not observe their realized values at $t = 1$. These net supply shocks are a standard feature of noisy rational expectations models, and can be interpreted as the result of trades which occur for life-cycle or liquidity reasons, or alternatively mistakes or lack of perfect information on the part of some investors (Admati, 1985, Van Nieuwerburgh and Veldkamp, 2010, Mondria, 2010). This assumption guarantees that prices do not perfectly reveal private information. If they did, differing posterior beliefs would not emerge in a setting with rational investors and common priors. Moreover, there would be no benefit to one investor in paying attention to private signals when all others have done so. Given the net supply assumptions, the following equilibrium conditions must hold: $\int X_{n0}^i di = 0$ and $\int X_{n1}^i di = a_n$ for all $n \in \{1, \ldots, N\}$.

We assume that prior distributions $\tilde{v}_{n0}^i$ and $\tilde{c}_{m0}$ are improper, characterized by infinite variance. As a result, any positive holdings of risky assets lead to infinite variance in wealth for $t = 2$, unless an agent has additional information with which to update her beliefs. Consequently, investors choose to hold only the risk-free asset at $t = 0$, a conclusion that would similarly arise in a setting with uniform proper priors, given our assumption of zero net supply at $t = 0$. The assumption of improper priors does not qualitatively alter the conclusions derived from the analysis,[3] but it does simplify the solution to the investors' attention allocation problem, which we outline below. This model can be interpreted as a limiting case where the precision of priors approaches zero.

At $t = 1$, investors observe noisy private signals about each fundamental factor v_n and c_m, which they then use to update their beliefs about each factor. Specifically, the signals for the asset-specific and the category-level factor are, respectively,

$$s_{vn}^i = v_n + \omega_{vn}^i \tag{3}$$

[3] Appendix C further discusses this assumption and relates it to the existent literature.

$$s^i_{cm} = c_m + \omega^i_{cm} \tag{4}$$

where the signal noises ω^i_{vn} and ω^i_{cm} are assumed to have Gaussian distributions with zero means and finite variances. The signal noises are independent of each other, because they are generated through separate tasks. Their variances depend on investors' attention allocations. Importantly, investors decide how much attention to allocate to each factor *before* observing the private signals. The relationship between attention and signal variance is given by

$$V\left(\omega^i_{vn}\right) = \frac{1}{p(\gamma^i_{vn})}, V\left(\omega^i_{cm}\right) = \frac{1}{p(\gamma^i_{cm})}$$

where $p : \mathbb{R} \to \mathbb{R}$ is a concave, strictly increasing function in C^2 and $\gamma^i_{vn}, \gamma^i_{cm} \in \mathbb{R}$ denote the attention paid to factors v_n and c_m, respectively.

Intuitively, when agents spend time learning about a given factor –or devote cognitive resources to it– they develop more precise signals about its true value. In addition, when deciding how much attention to allocate to each factor, investors must satisfy a resource constraint on the total amount of available attention. We express the attention constraint as

$$\sum_{n=1}^{N} \gamma^i_{vn} + \sum_{m\in\{A,B\}} \gamma^i_{cm} \leq \kappa \tag{5}$$

where κ is assumed equal for all investors $i \in [0, 1]$. If $p(\gamma) = C\gamma$, with $C > 0$, the relationship between signal precision (the variance's reciprocal) and attention is linear. This along with the linear attention constraint (5) lead to what Van Nieuwerburgh and Veldkamp (2010) refer to as a learning technology with additive precision.[4] We can interpret this as the continuous analog of a situation where investors learn by observing a sequence of independent draws of unitary variance, and the amount of attention they allocate to each factor is given by the amount of draws they observe for said factor (note that in this case, observing $\gamma \in \mathbb{N}$ draws of an asset would result in a variance of $\frac{1}{\gamma}$).

The assumption of improper priors implies that, upon observing signals (but before observing prices), investors' Bayesian updating process assigns full weight to the signal. This means that their revised beliefs about fundamental factors v_n and c_m will be normally distributed, with means given by their respective signal draws, s^i_{vn} and s^i_{cm}, and variances given by $\frac{1}{p(\gamma_{vn})}$ and $\frac{1}{p(\gamma_{cm})}$, respectively. Consequently, investor i's beliefs about D_n after observing signals are normally distributed as well, with a mean of $s^i_{vn} + s^i_{cm}$ and a variance of $\frac{1}{p(\gamma_{vm})} + \frac{1}{p(\gamma_{cm})}$, where $m = m(n)$. In turn, the covariance between post-signal beliefs for any two assets in category $m \in \{A, B\}$ is $\frac{1}{p(\gamma_{cm})}$. Beliefs about payoffs for assets belonging to different categories have a covariance of zero.

[4] In Van Nieuwerburgh and Veldkamp (2010), investors directly choose precision matrices, with the constraint expressed in terms of the difference between prior and posterior precision (in the case of additive precision). Our formulation is mathematically equivalent, given our assumption that the precision of prior beliefs is 0.

We write the vector of total signals for each asset as:

$$\boldsymbol{\zeta^i} \equiv (s^i_{v1} + s^i_{cA}, \ldots, s^i_{vN_A} + s^i_{cA}, s^i_{vN_A+1} + s^i_{cB}, \ldots, s^i_{vN} + s^i_{cB})^T$$

After observing her private signals, investor i's posterior beliefs about the payoff vector $\mathbf{D} \equiv (D_1, \ldots, D_n)^T$ follow a multivariate normal distribution, with mean equal to the realized ζ^i and a precision matrix equal to that of ζ^i, which we denote by Λ^i. Conditional on choosing attentions $\gamma^i_{vA}, \gamma^i_{vB}, \gamma^i_{cA}$ and γ^i_{cB}, the covariance matrix of ζ^i (the inverse of Λ^i) becomes:

$$\left(\Lambda^i\right)^{-1} = \begin{pmatrix} M^i_A & 0_{N_A \times N_B} \\ 0_{N_B \times N_A} & M^i_B \end{pmatrix} \tag{6}$$

where $0_{R \times C}$ denotes a $R \times C$ matrix filled with 0's. The matrices M^i_A and M^i_B are square matrices of dimensions N_A and N_B, respectively, which are given by:

$$M^i_A = \begin{pmatrix} \frac{1}{p(\gamma^i_{v1})} + \frac{1}{p(\gamma^i_{cA})} & \frac{1}{p(\gamma^i_{cA})} & \cdots & \frac{1}{p(\gamma^i_{cA})} \\ \frac{1}{p(\gamma^i_{cA})} & \frac{1}{p(\gamma^i_{v2})} + \frac{1}{p(\gamma^i_{cA})} & \cdots & \frac{1}{p(\gamma^i_{cA})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p(\gamma^i_{cA})} & \frac{1}{p(\gamma^i_{cA})} & \cdots & \frac{1}{p(\gamma^i_{vN_A})} + \frac{1}{p(\gamma^i_{cA})} \end{pmatrix} \tag{7}$$

and

$$M^i_B = \begin{pmatrix} \frac{1}{p(\gamma^i_{vN_A+1})} + \frac{1}{p(\gamma^i_{cB})} & \frac{1}{p(\gamma^i_{cB})} & \cdots & \frac{1}{p(\gamma^i_{cB})} \\ \frac{1}{p(\gamma^i_{cB})} & \frac{1}{p(\gamma^i_{vN_A+2})} + \frac{1}{p(\gamma^i_{cB})} & \cdots & \frac{1}{p(\gamma^i_{cB})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p(\gamma^i_{cB})} & \frac{1}{p(\gamma^i_{cB})} & \cdots & \frac{1}{p(\gamma^i_{vN})} + \frac{1}{p(\gamma^i_{cB})} \end{pmatrix} \tag{8}$$

We assume that all investors know the true joint distribution of the signals conditional on each possible attention allocation.

Finally, the timing of the model is as follows. In $t = 0$, all agents, endowed with a wealth of W_0 each, decide on a portfolio they wish to hold, based on their prior beliefs about payoffs.[5] Period $t = 1$ is divided into three stages. In the first stage, investors decide how much attention to pay to each signal. In the second stage, they observe noisy private signals, the precision of which is determined by the previously decided attention allocation. They then update their beliefs about payoffs D_n for each asset n. Net supply shocks realize. In the third stage, investors decide how much of each asset to hold and trade to adjust their portfolios in their desired allocations. Finally, at $t = 2$, the only event is the realization of all payoffs D_n (and 1 for the risk-free asset).

[5] We henceforth pay little attention to $t = 0$ because, as noted, investors hold only the riskless asset in equilibrium during this period. We include this period in the model so that investors' trade in $t = 1$ can be computed as the difference between their portfolio holdings in each period.

3.2 Market equilibrium

3.2.1 Definition of the trading stage equilibrium in $t = 1$

Let $\mathbf{X_t^i} \equiv (X_{1t}^i, \ldots, X_{Nt}^i)^T$, $\mathbf{P_t} \equiv (P_{1t}, \ldots, P_{Nt})^T$ and $\mathbf{a} \equiv (a_1, \ldots, a_N)^T$ be the vectors of intermediate asset holdings for investor i, prices and net supply for every asset, respectively, for a given period t. We define a market equilibrium in period t as a vector of prices $\mathbf{P_t}$ and a set of asset holdings $\mathbf{X_t^i}$, one for each investor $i \in [0, 1]$, such that:

- $\mathbf{X_t^i}$ maximizes investors i's expected utility given $\mathbf{P_t}$ among vectors that are feasible for i, satisfying equation (2) in $t = 0$ or equation (1) in $t = 1$.
- Aggregate net demand, given by the integral over i of $\mathbf{X_t^i}$, equals net supply (which equals the null vector if $t = 0$ and $\mathbf{a}$ if $t = 1$).

In this equilibrium, prices adjust so that players' net demand decisions, which are influenced by their beliefs about asset payoffs, are compatible, in aggregate, with the net supply of each asset n, given by a_n. An important implication of this is that, even though prices do not perfectly convey aggregate private information (due to investors' lack of knowledge about the actual net supply shocks), they still provide valuable insights into the signal realizations of every investor in the market. These signals, in turn, offer information about the true values of future payoffs $D_1, \ldots, D_N$. Aware of this dynamic, investors update their beliefs about payoffs based on observed asset prices. Similar to other noisy rational expectations equilibrium models (Admati, 1985), this creates a fixed-point problem, where equilibrium prices influence private beliefs (by transmitting information about the beliefs of other players) and are simultaneously determined by them.

We are not concerned with $\mathbf{P_0}$, since all that is important for the purpose of determining trade in the following period is the knowdedge that $\mathbf{X_0} = 0_{N\times 1}$. To solve for the equilibrium in $t = 1$, we focus on linear price functions where agents assume that:

$$\mathbf{P_1} = M_D\mathbf{D} + M_a\mathbf{a} \tag{9}$$

where M_D and M_a are matrices (and thus $M_D\mathbf{D}$ and $M_a\mathbf{a}$ are linear transformations of $\mathbf{D}$ and $\mathbf{a}$, respectively), which we furthermore assume to be invertible. The relative simplicity of deriving this equilibrium and the associated belief-updating process for investors contributes to its realism. By assuming less sophistication on the part of investors, the model better reflects real scenarios.

An implication of equation (9) is that $M_D^{-1}\mathbf{P_1}$ follows a normal distribution centered around $\mathbf{D}$, which investors interpret as a normally distributed signal. Consequently, after adjusting based on equilibrium prices, investors' beliefs about $\mathbf{D}$ still continue to be represented by a multivariate normal distribution. According to equation (1), this implies that the beliefs of investor i regarding W_2^i, given any fixed $\mathbf{X_1^i}$, also remain normally distributed after observing

prices.

3.2.2 Finding investors' optimal holdings at $t = 1$

In what follows, let $E^i(.)$ and $V^i(.)$ represent expectation and variance, respectively, of the beliefs of agent i for any uncertain outcome. At $t = 1$, having observed both signals and prices, investor i has a given amount of wealth W_1^i and her maximizes her expected utility, given by:

$$E^i\left(u(W_2^i)|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) = E^i\left(-e^{-W_2^i}|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) = -e^{-E^i\left(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) + \frac{V^i\left(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}\right)}{2}} \tag{10}$$

where W_2^i is determined according to equation (1) and the second equality follows from the normality of W_2^i. Maximizing expected utility is equivalent to minimizing:

$$g^i(\mathbf{X_1^i}) = -E^i\left(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) + \frac{V^i\left(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}\right)}{2}$$

Relying on equation (1), we can express the expectation of W_2^i conditional on $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$ as:

$$E^i\left(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) = -W_1^i - \sum_{n=1}^{N}\left(E^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) - P_{n1}\right)X_{n1}^i \tag{11}$$

We can likewise use equation (1) to write the variance of W_2^i, conditional on $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$ as:

$$V^i(W_2^i|\boldsymbol{\zeta^i}, \mathbf{P_1}) = \sum_{n=1}^{N}\left(X_{n1}^i\right)^2 V^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) + 2\sum_{n=1}^{N}\sum_{n'<n} X_{n1}^i X_{n'1}^i COV^i(D_n, D_{n'}|\boldsymbol{\zeta^i}, \mathbf{P_1}) \tag{12}$$

The partial derivative of $g^i(\mathbf{X_1})$ with respect to X_{n1}^i (that is, the demanded holdings of one asset, not the entire vector $\mathbf{X_1^i}$), is then given by:

$$\begin{aligned} &-\frac{\partial g^i}{\partial E^i(W_1^i)}\frac{\partial E^i(W_1^i)}{\partial X_{n1}^i}(.|\boldsymbol{\zeta^i}, \mathbf{P_1}) + \frac{\partial g^i}{\partial V^i(W_1^i)}\frac{\partial V^i(W_1^i)}{\partial X_{n1}^i}(.|\boldsymbol{\zeta^i}, \mathbf{P_1}) = \\ &-\left[E^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) - P_{n1}\right] + X_{n1}^i V^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) + \sum_{n'\neq n} X_{n'1}^i COV^i(D_n, D_{n'}|\boldsymbol{\zeta^i}, \mathbf{P_1}) \end{aligned} \tag{13}$$

Deriving again the right-hand side (RHS) of equation (13) with respect to X_{n1}^i yields $V^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1})$, whereas deriving the same expression with respect to $X_{n'1}^i, n' \neq n$ yields $COV^i(D_n, D_{n'}|\boldsymbol{\zeta^i}, \mathbf{P_1})$. The Hessian matrix for $g^i(.)$ is thus equal to the covariance matrix for investor i's beliefs about $(D_1, \ldots, D_N)$, conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$, which means that it is positive definite. This means that $g^i(.)$ is strictly convex, and therefore the second order conditions are such that the point which renders all (first order) partial derivatives equal to 0 in fact minimizes $g^i(.)$, as intended. If we impose that the RHS of equation (13) equals 0, then:

$$X_{n1}^i V^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) + \sum_{n'\neq n} X_{n'1}^i COV^i(D_n, D_{n'}|\boldsymbol{\zeta^i}, \mathbf{P_1}) = E^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) - P_{n1} \tag{14}$$

We can define $\Omega^i \equiv \left(V^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1})\right)^{-1}$, the precision matrix for investor i's beliefs for $\mathbf{D}$ conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$. We refer to the n-th row (or column) vector of the covariance matrix $(\Omega^i)^{-1} = V^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1})$ as $(\Omega^i_n)^{-1}$. Note that the n-th element in $(\Omega^i_n)^{-1}$ is $V^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1})$, while the n'-th element is $COV^i(D_n, D_{n'}|\boldsymbol{\zeta^i}, \mathbf{P_1})$ for all $n' \neq n$. We can thus rewrite equation (14) as:

$$(\Omega^i_n)^{-1}\mathbf{X_1^i} = E^i(D_n|\boldsymbol{\zeta^i}, \mathbf{P_1}) - P_{n1} \tag{15}$$

and obtain:

$$(\Omega^i)^{-1}\mathbf{X_1^i} = E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) - \mathbf{P_1} \tag{16}$$

From equation (16), we obtain the net demand function:

$$\mathbf{X^i} = \Omega^i \left[E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) - \mathbf{P_1}\right] \tag{17}$$

Note that since $X^i_{n0} = 0 \; \forall \; n \in \{1, \ldots, N\}, i \in [0, 1]$, investor i's trade vector in $t = 1$ is given by $\Delta\mathbf{X_1^i} \equiv \mathbf{X_1^i} - \mathbf{X_0^i} = \mathbf{X_1^i}$. Hence, equation (17) also describes equilibrium trade by investor i.

3.2.3 Equilibrium price vector

We use $\int$ to denote the Lebesgue integral and assume that $\int \zeta^i di$ is well-defined and equal to $\mathbf{D}$ (that is, the law of large numbers holds).[6]

The aggregate net demand vector at $t = 1$ is given by:

$$\begin{aligned} \int \mathbf{X_1^i} di = \int \Omega^i \left[E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) - \mathbf{P_1}\right] di = \\ \int \Omega^i E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) di - \left(\int \Omega^i di\right) \mathbf{P_1} \end{aligned} \tag{18}$$

In equilibrium, it must be equal to the vector of net supply shocks $\mathbf{a}$. We postulate that, if there exists an invertible matrix $\bar{\Lambda}$ such that $\int \Lambda^i di = \bar{\Lambda}$ and $\int \Lambda^i \zeta^i di = \bar{\Lambda}\mathbf{D}$, the following price function describes equilibria:

$$\mathbf{P_1} = \mathbf{D} - \left(\bar{\Lambda}\right)^{-1} \mathbf{a} \tag{19}$$

Since $E(\mathbf{P_1}) = \mathbf{D}$, the price vector itself constitutes an unbiased signal of $\mathbf{D}$, which investors use to update their beliefs. Note that its covariance matrix for $V(\mathbf{P_1}) = \left(\bar{\Lambda}\right)^{-1} \frac{1}{q_a} I_N \left(\bar{\Lambda}\right)^{-1} = \frac{1}{q_a}\left(\bar{\Lambda}\right)^{-2}$, where I_N is the identity matrix of size N, given the assumption that net supply shocks a_n are all mutually independent and of equal variance $\frac{1}{q_a}$. Moreover, $\mathbf{P_1}$ is normally distributed, since it is an affine transformation of $\mathbf{a}$. We then have that $\mathbf{P_1} \sim \mathcal{N}(\mathbf{D}, \frac{1}{q_a}\Lambda^{-2})$, which means that the precision matrix of $\mathbf{P_1}$ is $q_a\bar{\Lambda}^2$.

[6] In some cases, for the set of realizations of ζ^i for all investors $i \in [0, 1]$, the Lebesgue integral $\int \zeta^i di$ may not be well-defined. Alternatively, even if the integral is well-defined, the law of large numbers may fail to hold. These issues can be addressed by extending the standard product probability space in a manner that ensures that the probability of $\int \zeta^i di$ being undefined is 0, and that the law of large numbers is restored (see for example Judd, 1985 and Van Nieuwerburgh and Veldkamp, 2010).

Given that investor $i \in [0,1]$ arrives at the trading stage of $t = 1$ with prior beliefs distributed $\mathcal{N}\left(\boldsymbol{\zeta^i}, \Lambda^i\right)$ (after starting with improper priors and observing $\boldsymbol{\zeta^i}$), her posterior beliefs after observing prices are normally distributed, with mean:

$$E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) = (\Lambda^i + q_a\bar{\Lambda}^2)^{-1}[\Lambda^i\boldsymbol{\zeta^i} + q_a\bar{\Lambda}^2\mathbf{P_1}] \tag{20}$$

The equation above says that investors' posterior beliefs are a weighted average of their private signal realizations and prices. The precision of investor i's posterior beliefs is given by $\Omega_1^i = \Lambda^i + q_a\bar{\Lambda}^2$. Since $\Omega^i E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) = \Lambda^i\boldsymbol{\zeta^i} + q_a\bar{\Lambda}^2\mathbf{P_1}$, it yields:

$$\int \Omega^i E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1})di = \int \Lambda^i\zeta^i di + q_a\bar{\Lambda}^2\mathbf{P_1} = \bar{\Lambda}\mathbf{D} + q_a\bar{\Lambda}^2\mathbf{P_1}$$

Moreover, since $\int \Lambda^i di = \bar{\Lambda}$, we have that $\int \Omega^i di$ is well defined and equal to $\bar{\Lambda} + q_a\bar{\Lambda}^2$. Thus, the equilibrium condition that equation (18) must equal $\mathbf{a}$ can be re-written as:

$$\begin{gathered}\bar{\Lambda}\mathbf{D} + q_a\bar{\Lambda}\mathbf{P_1} - \left(\bar{\Lambda} + q_a\bar{\Lambda}^2\right)\mathbf{P_1} = \mathbf{a} \iff \\ \bar{\Lambda}\mathbf{D} - \bar{\Lambda}\mathbf{P_1} = \mathbf{a}\end{gathered}$$

Using the invertibility of $\bar{\Lambda}$, we obtain

$$\mathbf{P_1} = \mathbf{D} - \left(\bar{\Lambda}\right)^{-1}\mathbf{a}$$

which proves that the function in equation (19) describes equilibrium prices, as postulated. In all further sections, we assume that the trading stage equilibrium is of the type described above, and that this is anticipated by all investors.

3.2.4 Trade in equilibrium

Applying equation (16), we have that optimal holdings by investor i are given by

$$\begin{aligned}\mathbf{X_1^i} = (\Lambda^i + q_a\bar{\Lambda}^2)[E^i(\mathbf{D}|\boldsymbol{\zeta^i}, \mathbf{P_1}) - \mathbf{P_1}] = \Lambda^i\boldsymbol{\zeta^i} + q_a\bar{\Lambda}^2\mathbf{P_1} - (\Lambda^i + q_a\bar{\Lambda}^2)\mathbf{P_1} = \Lambda^i[\boldsymbol{\zeta^i} - \mathbf{P_1}] = \\ = \Lambda^i[\boldsymbol{\zeta^i} - \mathbf{D} + \bar{\Lambda}^{-1}\mathbf{a}]\end{aligned} \tag{21}$$

We define $\omega_n^i \equiv \omega_{vn}^i + \omega_{cm}^i$, with $m(n) = m$, and $\boldsymbol{\omega^i} \equiv (\omega_1^i, \ldots, \omega_N^i)^T$. This means that, from equation (21), we obtain

$$\mathbf{X_1^i} = \Lambda^i\left[\boldsymbol{\omega^i} + \bar{\Lambda}^{-1}\mathbf{a}\right] \tag{22}$$

In our empirical application, we are interested in trade volume. Note that its model counterpart is not $\mathbf{X_1^i}$, which counts sales of assets as negative trade, but the absolute value of $\mathbf{X_1^i}$. It is thus useful to derive results for $E\left(|X_{n1}^i|\right)$, expected absolute trade in asset n for some investor i. An implication of equation (22) is, if $\Lambda^i = \bar{\Lambda} = \Lambda$, then the expectation of each element n in vector $\left|\mathbf{X_1^i}\right|$ is increasing in Λ_{nn}, its corresponding element in the diagonal of the signal

precision matrix Λ. The precise relationship is given by

$$E(|X_{n1}^i|) = \sqrt{\frac{2\left(\Lambda_{nn} + \frac{1}{q_a}\right)}{\pi}} \tag{23}$$

Appendix D contains a proof of this result.

To understand the relationship between signal precision and trade, we turn to equations (22) and (23).[7] Equation (22) states that the *sign* of investor i's trade (i.e., whether she buys or sells) depends on the difference between the realization of her signal errors (ω^i) and on the equilibrium pricing errors ($\bar{\Lambda}^{-1}\mathbf{a} = \mathbf{D} - \mathbf{P_1}$ from equation (19)). Investor i's trade, in either direction, is hence increasing in the absolute value of the sum of these errors. Both effects constitute a channel through which signal and pricing errors affect investor i's trade (given by X_1^i). We refer to this channel as the *error channel*.

However, investor's trade (in either direction) is also increasing in the precision of her signals (captured by Λ^i), since a higher precision increases her propensity to trade, hereafter, the *precision channel*. Given the above channels, the relationship between signal precision and absolute trade is, at first glance, ambiguous. An investor with more precise signals is likely to disagree less with the market (the volatility of $\omega^i + \bar{\Lambda}^{-1}\mathbf{a}$ is reduced) but she is more prone to bet against it, given any fixed amount of disagreement. Additionally, in an equilibrium in which signal precisions are greater for all investors, the average precision (captured by $\bar{\Lambda}$) is greater, and, as a result, the absolute pricing errors ($\bar{\Lambda}^{-1}\mathbf{a}$) tend to be lower in absolute value. This also reduces the volatility of $\omega^i + \bar{\Lambda}^{-1}\mathbf{a}$, and thus pushes the relationship between signal precision and equilibrium absolute trade downwards. In turn, equation (23) indicates that in a symmetric equilibrium, the precision channel dominates, leading to a positive net effect on expected absolute trade.

A higher degree of confidence in one's disagreement with market prices, shared by all investors, remains consistent with rational behavior due to the presence of random net supply shocks. One might think that more precise signals imply more accurate aggregate private information at the market level, which would make investors more cautious about acting on their disagreement with prices. Indeed, higher signal precision means that prices themselves become a more reliable indicator of future payoffs, as suggested by equation (19) and discussed earlier. However, net supply shocks benefit investors on average because positive supply shocks undervalue assets (given equation (19)), making purchases profitable, while also allowing equilibria in which investors, in the aggregate, buy more assets than they sell.

Investors' trading strategy thus involves discerning whether discrepancies between signals and prices stem from signal imprecision (which would discourage trading) or from net supply shocks, which they can rationally exploit for profit. Crucially, net supply shocks are uninformative about the underlying assets' fundamentals, meaning that any price impacts they generate

[7] While equation (22) is written in matrix form, we can derive its intuition by interpreting it in scalar terms.

present opportunities for investors with favorable signal realizations to profit, rather than reasons to revise their beliefs about asset payoffs.

3.3 The attention allocation problem

Investor i's attention allocation problem in the first stage of $t = 1$ determines the precision matrix Λ of the signal vector $\boldsymbol{\zeta^i}$ that investor i receives in the second stage. Let $\widetilde{\boldsymbol{\gamma}}^{\boldsymbol{i}} \equiv (\gamma^i_{v1}, \ldots, \gamma^i_{vN_A}, \gamma^i_{cA}, \gamma^i_{vN_A+1}, \ldots, \gamma^i_{vN}, \gamma^i_{cB})^T$ denotes the vector of attentions chosen by investor i. We can then write the problem that investor i solves as follows:

$$max_{\widetilde{\boldsymbol{\gamma}}^{\boldsymbol{i}}} E^i \left(-e^{-W_2^i}\right)$$

$$\text{subject to}$$

$$\sum_{n=1}^{N_A} \gamma^i_{vn} + \gamma^i_{cA} + \sum_{n=N_A+1}^{N} \gamma^i_{vn} + \gamma^i_{cB} \leq \kappa$$

$$\gamma^i_{vn} \geq 0 \text{ for all } n \in \{1, \ldots, N\}$$

$$\gamma^i_{cA} \geq 0$$

$$\gamma^i_{cB} \geq 0$$

In the first stage of $t = 1$, investor i must choose $\boldsymbol{\gamma^i}$ so as to maximize her expected utility $E\left(-e^{-W_2^i}\right)$, knowing that her beliefs and behavior follow the timing described above.

Investor i's attention allocation determines her signal precision matrix Λ^i, as discussed in Section 3.1. In fact, this is the only channel through which it affects her expected utility. Similarly, the attention allocations of other players only affect investor i through their influence on the signal precision matrix, $\bar{\Lambda} = \int \Lambda^j dj$. This matrix is the same regardless of i's choice of Λ^i, because she is of measure 0. Thus, the problem stated above is equivalent to one of choosing Λ^i while taking $\bar{\Lambda}$ as given, and bearing in mind the equilibrium conditions for the trading stage of $t = 1$.

The following theorem shows that solving this problem is equivalent to maximizing a specific matrix determinant.

Theorem 1. *Investor i's attention allocation problem is equivalent to:*

$$max_{\boldsymbol{\gamma^i}} \left| I_N + \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2} \right| \tag{24}$$

$$\textit{subject to}$$

$$\sum_{n=1}^{N_A} \gamma^i_{vn} + \gamma^i_{cA} + \sum_{n=N_A+1}^{N} \gamma^i_{vn} + \gamma^i_{cB} \leq \kappa$$

$$\gamma^i_{vn} \geq 0 \ \forall \ n \in \{1, \ldots, N\}$$

$$\gamma^i_{cm} \geq 0 \ \forall \ m \in \{A, B\}$$

where $\boldsymbol{\gamma^i} = (\gamma^i_{v1}, \ldots, \gamma^i_{vN_A}, \gamma^i_{vN_A+1}, \ldots, \gamma^i_{vN}, \gamma^i_{cA}, \gamma^i_{cB})^T$ is the attention vector chosen by investor i, which yields a signal precision matrix of Λ^i, and $\bar{\Lambda}$ represents the signal precision matrix faced by all investors $j \in [0,1] \backslash \{i\}$ after having chosen $\bar{\boldsymbol{\gamma}}$.

Appendix E.1 first outlines and and then presents the full proof of Theorem 1.

Hereafter, we focus on equilibria where all investors allocate the same amount of attention to each asset-specific fundamental factor within a group m, that is $\gamma^i_{vn} = \gamma^i_{vn'} = \gamma^i_{vm}$ whenever $m(n) = m(n') = m$, for all $i \in [0,1]$. We refer to this property as *attention symmetry within a category*. Intuitively, focusing on this type of equilibria makes sense when we consider that investors are *ex-ante* identical (at the attention allocation stage); therefore, they have no incentives to allocate different levels of attention to distinct assets within the same category. Attention allocations that exhibit this property lead to signal covariance matrices where the diagonal of each block matrix $M^i_m, m \in \{A, B\}$ (recall equations (7) and (8)) is constant, with every element given by $\frac{1}{p(\gamma^i_{vm})} + \frac{1}{p(\gamma^i_{cm})}$.

One convenient aspect of considering equilibria with attention symmetry within a given category is that the dimensionality of the attention allocation problem is reduced. Within an equilibrium of this nature, investor i's optimal strategy implies setting $\gamma^i_{vn} = \gamma^i_{vn'} = \gamma^i_{vm}$ whenever $m(n) = m(n') = m$. As a result, her attention allocation problem is equivalent to one of choosing $\boldsymbol{\gamma^i} \equiv (\gamma^i_{vA}, \gamma^i_{cA}, \gamma^i_{vB}, \gamma^i_{cB})^T \in \mathbb{R}^4$ so as to maximize her expected utility. We thus write all following results in terms of investor i's reduced attention allocation vector $\boldsymbol{\gamma^i}$ and an analogously derived vector

$$\bar{\boldsymbol{\gamma}} \equiv (\bar{\gamma}_{vA}, \bar{\gamma}_{cA}, \bar{\gamma}_{vB}, \bar{\gamma}_{cB})^T \in \mathbb{R}^4$$

for a given investor.

The following proposition states that the set of equilibria we wish to focus on is not empty. It also states that the previously described covariance matrices within this set of equilibria are defined by $\bar{\Lambda}^{-1}$.

Proposition 1. *There exists an equilibrium in which all investors allocate the same attention allocation within a category. Furthermore, in such an equilibrium, the signal covariance matrix*

$\bar{\Lambda}^{-1}$ is a block diagonal matrix, whose diagonal blocks $\bar{M}_m$, $m \in \{A, B\}$ have constant diagonals where each element is given by $\frac{1}{p(\gamma_{vm})} + \frac{1}{p(\gamma_{vm})}$.

Internet Appendix presents the full proof of Proposition 1.

By imposing that all elements of $\bar{\Lambda}^{-1}$ are finite, we can simplify the problem further, as the following proposition details.

Proposition 2. *In an equilibrium with attention symmetry within categories, if $p(\bar{\gamma}) > 0$ for all $\bar{\gamma} \in \{\bar{\gamma}_{vA}, \bar{\gamma}_{cA}, \bar{\gamma}_{vB}, \bar{\gamma}_{cB}\}$, the following equality holds for the determinant in Theorem 1:*

$$\left|I_N + q_a \Lambda^i \bar{\Lambda}^{-2}\right| = \left(1 + \frac{p(\gamma^i_{vA})}{q_a \left[p(\bar{\gamma}_{vA})\right]^2}\right)^{N_A - 1} \left(1 + \left(\frac{p(\gamma^i_{vA})p(\gamma^i_{cA})}{p(\gamma^i_{cA}) + N_A p(\gamma^i_{vA})}\right) \frac{1}{q_a} \left(\frac{p(\bar{\gamma}_{cA}) + N_A p(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2\right)$$
$$\left(1 + \frac{p(\gamma^i_{vB})}{q_a \left[p(\bar{\gamma}_{vB})\right]^2}\right)^{N_B - 1} \left(1 + \left(\frac{p(\gamma^i_{vB})p(\gamma^i_{cB})}{p(\gamma^i_{cB}) + N_B p(\gamma^i_{vB})}\right) \frac{1}{q_a} \left(\frac{p(\bar{\gamma}_{cB}) + N_B p(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2\right) \tag{25}$$

Appendix E.2 first outlines and and then presents the full proof of Proposition 2.

Maximizing the expression on the RHS of equation (25) is equivalent to maximizing its logarithm. Proposition 2 thus allows us to define, for any attention allocation $\bar{\boldsymbol{\gamma}} \equiv (\bar{\gamma}_{vA}, \bar{\gamma}_{cA}, \bar{\gamma}_{vB}, \bar{\gamma}_{cB})^T$ chosen by players $j \in [0,1]\backslash\{i\}$, a function $f_{\bar{\boldsymbol{\gamma}}} : \mathbb{R}^4 \to \mathbb{R}$, representing (the logarithm of) expected utility in the attention allocation stage for agent $i \in [0,1]$ depending on her attention allocation vector $\boldsymbol{\gamma^i} \equiv (\gamma^i_{vA}, \gamma^i_{cA}, \gamma^i_{vA}, \gamma^i_{cA})^T$. This function is given by:

$$f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i}) = (N_A - 1)\, ln\left(1 + \frac{p(\gamma^i_{vA})}{q_a \left[p(\bar{\gamma}_{vA})\right]^2}\right) +$$
$$ln\left(1 + \left(\frac{p(\gamma^i_{vA})p(\gamma^i_{cA})}{p(\gamma^i_{cA}) + N_A p(\gamma^i_{vA})}\right) \frac{1}{q_a} \left(\frac{p(\bar{\gamma}_{cA}) + N_A p(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2\right) + (N_B - 1)\, ln\left(1 + \frac{p(\gamma^i_{vB})}{q_a \left[p(\bar{\gamma}_{vB})\right]^2}\right) +$$
$$ln\left(1 + \left(\frac{p(\gamma^i_{vB})p(\gamma^i_{cB})}{p(\gamma^i_{cB}) + N_B p(\gamma^i_{vB})}\right) \frac{1}{q_a} \left(\frac{p(\bar{\gamma}_{cB}) + N_B p(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2\right) \tag{26}$$

The problem facing the investor becomes:

$$max_{\boldsymbol{\gamma^i}}\ f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i})$$
$$\text{subject to}$$
$$N_A \gamma^i_{vA} + \gamma^i_{cA} + N_B \gamma^i_{vB} + \gamma^i_{cB} \leq \kappa \tag{27}$$
$$\gamma^i_{vA} \geq 0 \tag{28}$$
$$\gamma^i_{vB} \geq 0 \tag{29}$$
$$\gamma^i_{cA} \geq 0 \tag{30}$$
$$\gamma^i_{cB} \geq 0 \tag{31}$$

We consider equilibria with interior solutions, so restrictions (28) through (31) are not binding. Given that $f(.)$ is strictly increasing in all arguments, it follows that restriction (27) is satisfied

with equality. We can thus obtain solutions using the Lagrangian:

$$\mathcal{L} = f(\gamma^i_{vA}, \gamma^i_{vB}, \gamma^i_{cA}, \gamma^i_{cB}) + \lambda(\kappa - N_A\gamma^i_{vA} - N_B\gamma^i_{vB} - \gamma^i_{cA} - \gamma^i_{cB})$$

The first-order conditions follow:

$$\frac{\frac{(N_A-1)p'(\gamma^i_{vA})}{q_a[p(\bar{\gamma}_{vA})]^2}}{\left(1+\frac{p(\gamma^i_{vA})}{q_a[p(\bar{\gamma}_{vA})]^2}\right)} + \frac{\left(\frac{p'(\gamma^i_{vA})[p(\gamma^i_{cA})]^2}{\left[p(\gamma^i_{cA})+N_Ap(\gamma^i_{vA})\right]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA})+N_Ap(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2}{1+\left(\frac{p(\gamma^i_{vA})p(\gamma^i_{cA})}{p(\gamma^i_{cA})+N_Ap(\gamma^i_{vA})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA})+N_Ap(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2} = N_A\lambda \tag{32}$$

$$\frac{\left(\frac{N_Ap'(\gamma^i_{cA})[p(\gamma^i_{vA})]^2}{\left[p(\gamma^i_{cA})+N_Ap(\gamma^i_{vA})\right]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA})+N_Ap(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2}{1+\left(\frac{p(\gamma^i_{vA})p(\gamma^i_{cA})}{p(\gamma^i_{cA})+N_Ap(\gamma^i_{vA})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA})+N_Ap(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2} = \lambda \tag{33}$$

$$\frac{\frac{(N_B-1)p'(\gamma^i_{vB})}{q_a[p(\bar{\gamma}_{vB})]^2}}{\left(1+\frac{p(\gamma^i_{vB})}{q_a[p(\bar{\gamma}_{vB})]^2}\right)} + \frac{\left(\frac{p'(\gamma^i_{vB})[p(\gamma^i_{cB})]^2}{\left[p(\gamma^i_{cB})+N_Bp(\gamma^i_{vB})\right]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB})+N_Bp(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2}{1+\left(\frac{p(\gamma^i_{vB})p(\gamma^i_{cB})}{p(\gamma^i_{cB})+N_Bp(\gamma^i_{vB})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB})+N_Bp(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2} = N_B\lambda \tag{34}$$

$$\frac{\left(\frac{N_Bp'(\gamma^i_{cB})[p(\gamma^i_{vB})]^2}{\left[p(\gamma^i_{cB})+N_Bp(\gamma^i_{vB})\right]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB})+N_Bp(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2}{1+\left(\frac{p(\gamma^i_{vB})p(\gamma^i_{cB})}{p(\gamma^i_{cB})+N_Bp(\gamma^i_{vB})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB})+N_Bp(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2} = \lambda \tag{35}$$

3.3.1 The equilibrium of the attention allocation problem

The following proposition states that there is only one equilibrium within the subset of equilibria that we focus on, and that this equilibrium is symmetric.

Proposition 3. *There is a unique equilibrium with attention symmetry within a given category in which all investors select interior solutions. This equilibrium is symmetric, i.e.* $\boldsymbol{\gamma^i} = \boldsymbol{\gamma^j}$ *for all* $i, j \in [0, 1]$.

Appendix E.3 details the Proof of Proposition 3.

Since the equilibrium is symmetric, for the first-order conditions previously obtained, we have that $\boldsymbol{\gamma^i} = \bar{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \equiv (\gamma_{vA}, \gamma_{cA}, \gamma_{vB}, \gamma_{cB})^T$. Equations (32)-(35) then simplify to :

$$\frac{(N_A-1)\,p'(\gamma_{vA})}{q_a\left[p(\gamma_{vA})\right]^2 + p(\gamma_{vA})} + \frac{p'(\gamma_{vA})p(\gamma_{cA})}{p(\gamma_{vA})\left[q_ap(\gamma_{vA})p(\gamma_{cA}) + p(\gamma_{cA}) + N_Ap(\gamma_{vA})\right]} = N_A\lambda \tag{36}$$

$$\frac{N_Ap'(\gamma_{cA})p(\gamma_{vA})}{p(\gamma_{cA}\left[q_ap(\gamma_{vA})p(\gamma_{cA}) + p(\gamma_{cA}) + N_Ap(\gamma_{vA})\right])} = \lambda \tag{37}$$

$$\frac{(N_B-1)\,p'(\gamma_{vB})}{q_a\left[p(\gamma_{vB})\right]^2+p(\gamma_{vB})}+\frac{p'(\gamma_{vB})p(\gamma_{cB})}{p(\gamma_{vB})\left[q_a p(\gamma_{vB})p(\gamma_{cB})+p(\gamma_{cB})+N_B p(\gamma_{vB})\right]}=N_B\lambda \tag{38}$$

$$\frac{N_B p'(\gamma_{cB})p(\gamma_{vB})}{p(\gamma_{cB}\left[q_a p(\gamma_{vB})p(\gamma_{cB})+p(\gamma_{cB})+N_B p(\gamma_{vB})\right])}=\lambda \tag{39}$$

The following proposition characterizes the equilibrium of the attention allocation problem.

Proposition 4. *For the interior equilibrium solution to the attention allocation problem, the following results hold:*

1. *There is category learning, that is,* $\gamma_{cm} > \gamma_{vm}$ *for* $m \in \{A, B\}$.
2. *If* $N_A > N_B$, *then* $\gamma_{cA} > \gamma_{cB}$.

If $p(.)$ *is a linear function, we can also show that:*

3. *If* $N_A > N_B$, *then* $\gamma_{vA} > \gamma_{vB}$.
4. *If* $N_A > N_B$, *expected trade is higher for assets in* A *than for assets in* B.

Appendix E.4 details the proof of Proposition 4.

The first result in Proposition 4 represents the core finding of our theoretical model. It demonstrates that attention-constrained investors allocate more attention to category-level factors than to asset-specific factors, leading to what we refer to as *category-learning behavior*. Intuitively, attention-constrained investors have incentives to focus their attention on signals that provide the greatest amount of information about future payoffs. Since categorical signals contain information relevant to a group of assets, whereas asset-specific signals are limited to individual assets, investors prefer to prioritize the former when they are attention-constrained. The second result in Proposition 4 complements the first by showing that attention-constrained investors devote more attention to the categorical factors of the largest groups. The intuition behind this result is that the informational benefit of category-level signals is higher for more central categories—those containing a larger number of assets—thus leading investors to concentrate their attention on these signals.

Result 3 emerges from the previous two results: investors allocate more attention to asset-specific factors when those assets belong to more central categories. The reason for this is that there are strategic complementarities between category-level and asset-specific attention: the incentives to pay attention to an asset-specific fundamental increase with the amount of attention paid to its category-level fundamentals. To see this, consider the precision of an investor's beliefs about any given asset n in category m as $\frac{1}{\frac{1}{p(\gamma_{vm})}+\frac{1}{p(\gamma_{cm})}}=\frac{p(\gamma_{vm})p(\gamma_{cm})}{p(\gamma_{vm})+p(\gamma_{cm})}$, which is the harmonic mean of the the precision of her beliefs about v_n and c_m. The harmonic mean has positive cross-derivatives, meaning that there are strategic complementarities between category-level and asset-specific attention.

Intuitively, for an investor to have sufficient confidence to make a bet against the market, she must be well-informed about both firm-level and category-level fundamentals. Therefore, when an investor chooses to be more informed about a particular category, paying attention to category-level fundamentals can lead to significant expected profits. Conversely, if a category is deemed less important and receives little attention, the investor understands that additional information at the asset level will have minimal impact on her profits. This effect does not contradict the concavity of the objective function (25), which implies diminishing marginal returns to attention devoted to a specific asset. Instead, result 3 arises from the complementarity between category-level and asset-specific attention.[8]

Finally, result 4 demonstrates that investors trade more frequently in assets they prioritize in their optimal allocation strategy. This occurs because, when $N_A > N_B$, $\gamma_{cA} > \gamma_{cB}$ and $\gamma_{vA} > \gamma_{vB}$ hold, as shown in results 2 and 3. As investor i allocates more attention to both categorical and asset-specific fundamental factors for assets belonging to the larger category, the *total* signal precision increases for assets in category A. Consequently, investor's posterior beliefs about the payoffs of assets in group A become more precise during the second stage at $t = 1$, leading to more accurate final beliefs after observing prices. Section 3.2.4 demonstrates that there are two main channels that affect differently the relationship between signal precision and absolute trade: the error and the precision channel. What result 4 in Proposition 4 shows is that the net result of increasing precision (due to higher categorical and asset-specific attention) on trade is positive: assets in the larger category are traded more, because investors are more willing to bet against the market in the event of disagreeing with equilibrium prices.

4 Data

In Section 3, we develop a model where attention-constrained investors decide how much attention to allocate between categorical and asset-specific fundamental factors. Our findings reveal that limited attention results in category-learning behavior, meaning that investors prioritize category-level factors over asset-specific ones. Building on this, we now propose a novel measure of category learning, which quantifies the extent to which two or more assets are collectively perceived as similar by users on a social network. The underlying intuition of our measure is that when users mention multiple assets together, they direct attention to them. Co-mentioning indicates that users likely perceive these assets as sharing common characteristics, or categorical factors. Next, we utilize this measure to empirically investigate the evolution of category learning and its relationship with a significant emerging trend in the financial industry: the rise of artificial intelligence.

The social network we consider is X (former Twitter). To extract users' collective perceptions of a stock's related peers, we apply a co-search algorithm to approximately 22.5 million posts

[8]However, the presence of diminishing marginal returns to attention does somewhat blunt this effect; without it, interior solutions would not be possible, and investors might entirely ignore smaller categories or, at the very least, the asset-specific fundamentals within those categories.

generated between January 2015 and December 2022, utilizing the X REST API.[9] Note that users in the empirical analysis approximate our investors in the theoretical model; therefore, we are implicitly considering retail investors (Li et al., 2016, Shen et al., 2019, Da et al., 2011). The API provides us with various data for each post, including its timestamp (in GMT), text, and the list of all cashtags mentioned. Each observation in our sample represents a time-stamped (hourly) post from a unique user that contains at least one cashtag associated with stocks in the S&P 500 index at time t (for example, \$AAPL, \$GS, \$DD).[10] From this universe of posts, we construct a *network of cashtags* that captures the extent to which different assets are mentioned together on X at a given point in time (an hour and a date); therefore, the cashtags building the network vary across periods.

We create an undirected network at each point in time, adding a node for each cashtag in the posts we download, regardless of whether the assets are S&P 500 constituents. This approach ensures that the network includes nodes representing S&P 500 companies as well as other assets outside the index.[11] Consider pairs of nodes in the network (e.g., \$AAPL and \$ETH) at a given point in time t. For each pair of nodes at time t, we count the number of posts that mention both cashtags. If this number is greater than zero, we add an edge between the two nodes and assign a weight to this edge equal to the number of co-mentions. If the count is zero, no edge is added.[12] Intuitively, our weight metric measures the degree to which two or more stocks are associated together in the collective perceptions of X users.

Figure 1 shows the network generated using the algorithm described earlier. This network focuses on posts posted within a single hour (August 25, 2021, from 13:00:00 to 14:00:00 EDT). To reduce clutter, we limit the links between stocks to posts with non-neutral sentiment. For sentiment classification, we apply FinBERT, a pre-trained BERT model fine-tuned for financial sentiment analysis (Huang et al., 2023). FinBERT assigns one of three possible sentiments—*positive*, *negative*, or *neutral*—to each X post written in the period analyzed and containing at least one S&P 500 cashtag.[13] We use only X posts classified as positive or neg-

[9] We are grateful to X for granting us access to the X Academic API for this research. While downloading the data, we exclude X posts created solely for promotional purposes or ads.

[10] It is important to note that searching for posts containing the tickers of S&P 500 stocks does not limit our study to these stocks alone, as each post may also include cashtags from assets not necessarily associated with S&P 500 companies, such as cryptocurrencies (e.g., \$BTC, \$ETH).

[11] Constituent information for the S&P 500 index is sourced from Refinitiv Eikon: https://eikon.refinitiv.com/.

[12] We also tested an alternative approach for constructing the network, using the positive pointwise mutual information (PPMI) metric as edge weights. PPMI quantifies the ratio between the probability of a post mentioning both cashtags simultaneously and the expected probability of this co-occurrence if the mentions were independent. It is defined as follows:

$$PPMI(X,Y) = \max\left(\log_2\left(\frac{\#posts_{X,Y} \cdot \#total\ posts}{\#posts_X \cdot \#posts_Y}\right), 0\right) \tag{40}$$

where X and Y represent the two given cashtags, $\#posts_{X,Y}$ corresponds to the number of posts mentioning both cashtags X and Y simultaneously, $\#posts_X$ ($\#posts_Y$) represents the number of posts mentioning cashtag X (cashtag Y), and $\#total\ posts$ represents the total number of posts being analyzed. The to-be presented results are qualitatively consistent to the ones we obtain when relying on the PPMI to build the network of cashtags.

[13] We also compare alternative methods, such as dictionary-based approaches and other supervised learn-

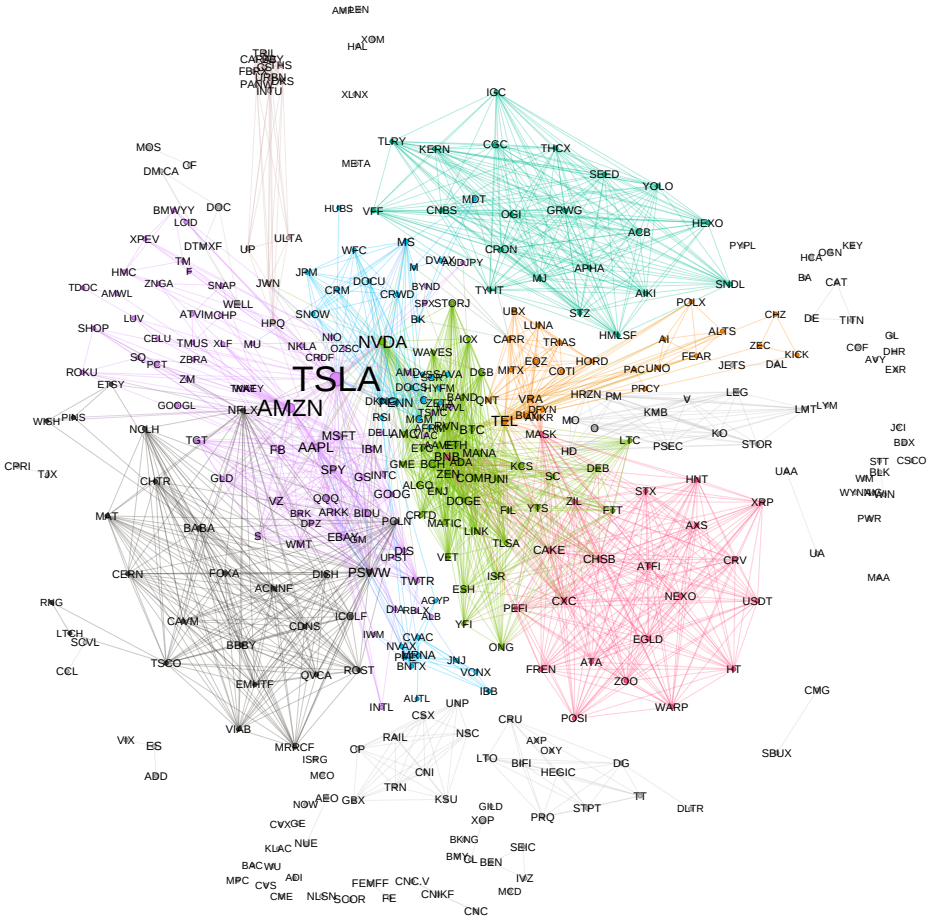
ative to construct three networks of cashtags: the positive, the negative, and the non-neutral networks. The inclusion of sentiment information is grounded in prior research demonstrating its predictive value for financial time series (Mohammad, 2016).

We construct the network in Figure 1 from 472 X posts, comprising 359 nodes and 2,106 edges.[14] The figure reveals a structured organization that appears to be non-random, with distinct communities identifiable, many corresponding to specific economic sectors. A large community, shown in violet, includes cashtags of technology-related companies (e.g., $AAPL, $AMZN, $TSLA), while another in light green primarily contains cryptocurrency-related cashtags (e.g., $BTC, $ETH, $DOGE). Other communities are also similar in terms of their main activities: for instance, a group positioned at the top of the plot includes companies associated with cannabis production and commercialization (e.g., $THCX, $CRONOS, $HEXO), and another community, highlighted in light blue, includes bank-related companies (e.g., $WFC, $JPM, $BK).

ing techniques, for sentiment classification. FinBERT provides classifications most consistent with the actual sentiments expressed in the posts. Additionally, the results are robust across various network constructions.

[14]For comparison, Figure A1 in the appendix displays the network obtained when all X posts are considered, regardless of sentiment, over the same period. This network is constructed from 2,185 X posts and consists of 1,468 nodes and 19,961 edges. As expected, the network in Figure A1 is significantly denser than the one in Figure 1.

Figure 1: Network of cashtags having a non-neutral valence



Notes: We construct this network of cashtags with the X posts containing an S&P 500 ticker classified as positive by FinBERT, published from August 25, 2021, 13:00:00 EDT to August 25, 2021, 14:00:00 EDT. Node sizes correspond to their PageRank and colors indicate communities identified using the Louvain method (Blondel et al., 2008).

We now describe how we construct our proxies for category-level and asset specific attention. To measure the attention given to categorical factors (γ_{cm}) and asset-specific factors (γ_{vm}) in the model, we utilize the network of cashtags on X at each point in time t. First, we identify the communities present in a network of cashtags at time t using the Louvain method for community detection (Blondel et al., 2008). This way, these identified communities serve as the empirical analogs of the theoretical categories m in the model. While the model simplifies the setup by including only two categories (A and B), this does not limit the generality of our results. To measure the number of assets within a category (N_m), we count the number of X posts in a given community at period t, thus turning the number of assets per community into N_{mt}, as communities evolve over time.

Our proxy for *categorical attention* (γ_{cm} in the model) corresponds to the number of X posts that mention at least two assets within the same community. Specifically, for each identified community m at time t, we count the posts mentioning more than two assets from that community, resulting in the measure γ_{cmt}. It is important to emphasize two underlying assumptions

in this framework. First, we assume that the number of X posts related to an asset reflects the level of attention users, and by extension investors, pay to that asset, approximating the degree of attention directed at the asset. Second, we assume that posts mentioning two or more cashtags within a community imply that users perceive these assets as similar. If users perceive assets as similar, it is likely that these assets share some categorical (common) factors, which affect the group of assets perceived as similar by X users. Recall that to identify these communities, the Louvain algorithm uses the weighted network of cashtags at each point in time.

We calculate the *total attention* given to community m at period t by counting the total number of X posts mentioning any asset within the community at time t. Finally, the difference between the total number of posts mentioning a community (reflecting overall attention) and our categorical attention measure (γ_{cmt}) represents the attention directed toward individual assets in the community. The theoretical model analog of this is $N_m\gamma_{vm}$, which captures asset-specific attention. We compute our proxy for γ_{vmt}, or average asset-specific attention, by dividing this difference by the total number of X posts mentioning the community at period t. This proxy, which evolves over time, represents the average attention paid to a single asset-specific factor within community m at time t.

Finally, to obtain intraday trade volume data, we rely on the Intraday Historical Data API provided by EOD Historical Data.[15] This API allows us to download data for selected assets at intervals of 1 minute, 5 minutes, and 1 hour. We download the hourly volume data for all stocks included as S&P 500 constituents at any given time, as well as hourly data on the volume of the S&P 500 index ETF.

5 Validation analysis

We now use our measure for categorical attention to empirically examine the evolution of category learning behavior and its relationship with key features in the financial industry, such as artificial intelligence. However, a natural question precedes: How do we know that this measure effectively captures category learning? In spite of the fact that our measure of category learning is based on a stylized model under frictions in the agent's learning process, we cannot offer a definite answer to this question. This is because learning dynamics and agents' attention constraints are not directly observable.

To indirectly address this question, we develop a set of testable implications, which we derive in Proposition 4, to investigate the behavior of our proposed measure of category learning under different scenarios, which we subsequently test with our data and using our proxies for categorical and asset-specific attention. These testable implications connect with previous literature offering potential explanations for the *existence* of category learning. This analysis can hence provide valuable insights into whether our proposed measure is able to capture

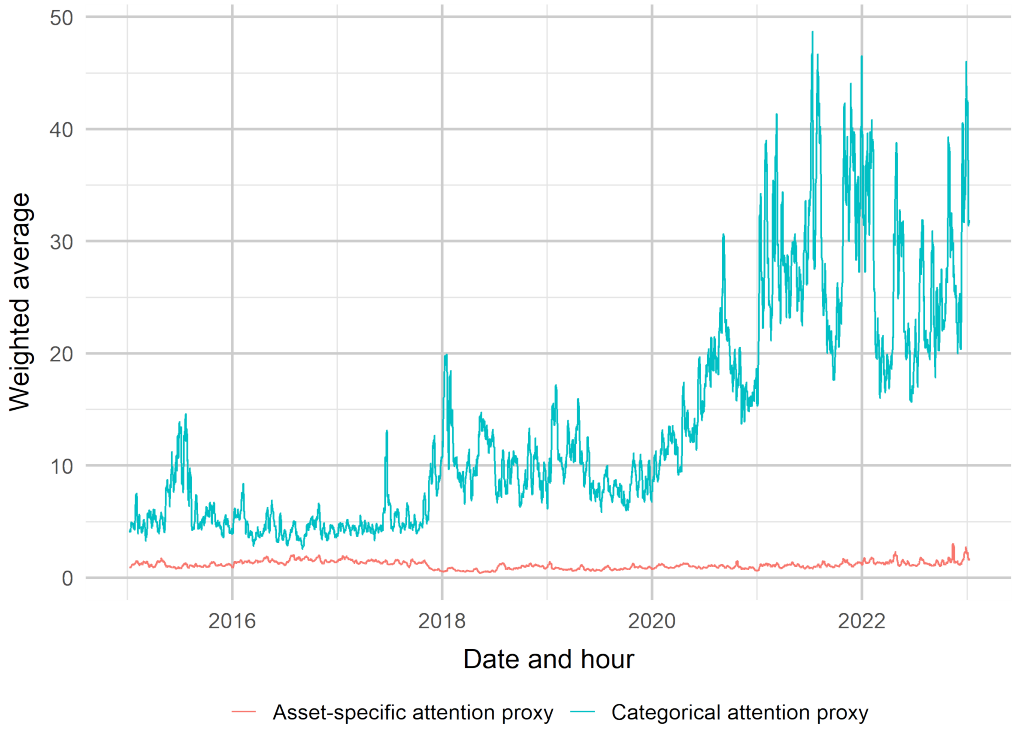
[15] https://eodhd.com/financial-apis/intraday-historical-data-api/.

category learning by generating results that are consistent with actual economic outcomes. We now test each of the implications in Proposition 4.

Result 1. Category-level attention is greater than asset-specific attention

To test this result, Figure 2 compares the evolution of our proxies for category-level and asset-specific attention, on average. The figure shows that our measure of category learning consistently exceeds the weighted average attention given to asset-specific factors, aligning with our theoretical model. Intuitively, attention-constrained investors focus more on signals that provide the greatest amount of information about future payoffs. Since categorical signals contain information about a group of assets, while asset-specific signals are limited to individual stocks, investors prefer to prioritize categorical signals when they are attention-constrained.

Figure 2: Proxies for categorical and asset-specific attention



Note: The figure shows the evolution of our proxies for categorical and asset-specific attention at the date-hour level. Each value represents the weighted weekly rolling average of X posts mentioning either two or more assets (for categorical attention) or a single asset (for asset-specific attention) across communities. These values are weighted by the number of assets in each community, and the measures are aggregated to produce rolling weekly averages weighted by assets within each period.[16]

The evidence in Figure 2 also aligns with Peng and Xiong (2006), who show that category-learning behavior (investors' tendency to process more market- and sector-level information) emerges in a model where a representative investor optimally makes consumption decisions based on beliefs about fundamental factors. In equilibrium, the investor allocates attention

[16]If an asset is not mentioned during a specific date-hour, it is assigned to its own community, the weight of which is set to zero.

across fundamental factors to minimize total portfolio uncertainty, while asset prices fluctuate as the investor updates beliefs based on processed information. In this way, the investor's attention allocation influences asset-price dynamics. Notably, the authors demonstrate that category-learning behavior, combined with overreaction to processed information (due to overconfidence bias), can explain certain features of asset-return co-movement observed in the data, specifically why return correlations between firms often exceed their fundamental correlations.

Result 2. Category-level attention is greater for larger categories

Result 2 shows that categorical attention is greater for larger categories. To test this theoretical result, we conduct a panel regression analysis. First, we perform the analysis at the community level, where each observation corresponds to a combination of community m and time (date-hour) t. The first model specification at community level writes as follows,

$$\gamma_{c,mt} = \alpha + \beta N_{mt} + \theta_1 N_{posts,mt} + \theta_2 N^2_{posts,mt} + F_{t-1}\delta + \epsilon_{mt} \tag{41}$$

As previously explained, we measure $\gamma_{c,mt}$, which is the empirical counterpart of the categorical attention (γ_{cm} in the model), by the number of X posts that mention at least two assets within the same community. To quantify community size N_{mt}, which is our independent variable of interest, we use two proxies: absolute and relative community size. Absolute community size refers to the number of assets within a community. Relative community size corresponds to the ratio between absolute community size and the total number of assets in the database for the given period. As control variable, we include the number of X posts per community at a point in time, which we refer as $N_{posts,mt}$.

As additional controls in equation (41), we incorporate a set of time-varying common factors, denoted by matrix F_t, which includes the VIX and the end-of-day implied correlation index (COR3M). The VIX captures market expectations of short-term volatility, as implied by the prices of a set of S&P 500 index options with near-term expiration dates. The COR3M index measures the expected average correlation between stock prices of S&P 500 components, using option prices from the S&P 500 index and its 50 largest components by market capitalization. Additionally, we include hour and day-of-week fixed effects.

Second, when observations correspond to a combination of asset i, community m and time t, the model specification becomes

$$\gamma_{c,mt} = \alpha + \beta N_{mt} + \theta_1 N_{posts,mt} + \theta_2 N^2_{posts,mt} + F_{t-1}\delta + \lambda_i + \epsilon_{imt} \tag{42}$$

where λ_i denotes asset-fixed effects.

There are two important points to note regarding the model specifications. First, the proxies for categorical and asset-specific attention in equations (41) and (42) are always measured at the community level. Second, even when the regression is run at the asset level, the number of posts, $N_{posts,mt}$, remains measured at the community level. This is because the dependent

variable is also at the community level. Therefore, including $N_{posts,mt}$ on the right-hand side serves as a normalization factor for category-level attention at the community level.

Table 1: Estimation results on categorical attention and community size. Assessing result 2

	Dependent variable: Categorical attention			
	Community level		Ticker level	
	(1)	(2)	(3)	(4)
Abs. community size	0.751***		4.447***	
	(0.001)		(0.003)	
Rel. community size		0.604***		3.361***
		(0.001)		(0.003)
Observations	6 180 640	6 180 640	7 097 727	7 097 727
Hour FE	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Ticker FE	No	No	Yes	Yes
R^2	0.828	0.822	0.912	0.904
Adjusted R^2	−0.062	−0.098	0.375	0.314

Note: This table presents the estimates from equations (41) and (42). Abs. and rel. stand for absolute and relative, respectively. The control variables in the regressions include the number of posts mentioning at least one asset within the community (in levels and squared), the first lag of the VIX, and the first lag of the COR3M. Estimates include hour, day-of-week and and ticker fixed effects (the latter when corresponding). The table excludes the control coefficients and intercept from all regressions. We standardize the community size variables. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001

The estimates in Table 1 support our theoretical result 2: users are more likely to associate two or more assets together when these assets belong to a larger community. These communities reflect the collective perceptions of X users regarding the *apparent* similarity among assets within the community. In the model, an informational mechanism drives this result: attention-constrained investors focus on signals that provide the greatest amount of information about future payoffs. Since categorical signals provide information about a group of assets, while asset-specific signals are limited to individual assets, investors prioritize categorical signals when they are attention-constrained. Moreover, larger communities —those containing more assets— offer more informational content, increasing the benefit of focusing on these signals. This explains why category-learning behavior is stronger for the categorical factors of larger groups, consistent with the empirical results in Table 1.

Result 3. Asset-specific attention is greater for larger categories

We now evaluate the empirical evidence for theoretical result 3. As in the previous analysis, we conduct panel regressions at both the community and the asset levels. The dependent variable, $\bar{\gamma}_{v,mt}$, represents the average attention directed to a single asset-specific factor within community m. We quantify this by calculating the average number of X posts mentioning a

given asset within a community. The model specifications for the community and asset levels, respectively, are as follows:

$$\bar{\gamma}_{v,mt} = \alpha + \beta N_{mt} + F_{t-1}\delta + \epsilon_{mt} \tag{43}$$

$$\bar{\gamma}_{v,mt} = \alpha + \beta N_{mt} + F_{t-1}\delta + \lambda_i + \epsilon_{imt} \tag{44}$$

Note that model specifications (43) and (44) do not include as control variables $N_{posts,mt}$ and $N^2_{posts,mt}$ given the way we construct the dependent variable. Moreover, it is a measure of the average attention directed to a single asset-specific factor within community m. Table 2 presents the results.

Table 2: Estimation results on asset-specific attention proxy and community size. Assessing result 3

	Dependent variable: Asset-specific attention			
	Community level		Ticker level	
	(1)	(2)	(3)	(4)
Abs. community size	4.325***		41.105***	
	(0.040)		(0.907)	
Rel. community size		5.259***		20.070***
		(0.040)		(0.921)
Observations	1 220 869	1 220 869	2 137 956	2 137 956
Hour FE	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Ticker FE	No	No	Yes	Yes
R^2	0.009	0.014	0.004	0.003

Note: This table presents the estimates from equations (43) and (44). Abs. and rel. stand for absolute and relative, respectively. The control variables in the regressions include the first lag of the VIX and the first lag of the COR3M. Estimates include hour, day-of-week and and ticker fixed effects (the latter when corresponding). The table excludes the control coefficients and intercept from all regressions. We standardize the community size variables. We drop from these regressions those communities consisting of a single asset with 0 associated X posts. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

The estimates in Table 2 also confirm theoretical result 3: users tend to give more asset-specific attention to assets in larger communities. Intuitively, larger communities are mentioned more frequently, increasing the likelihood that a user (our proxy for investors) will focus on a single asset within that community.

In the theoretical model, strategic complementarities between category-level and asset-specific attention explain this result. The incentives to focus on an asset-specific fundamental grow with the attention paid to its category-level fundamentals. Consequently, investors allocate

more attention to asset-specific factors when those assets belong to more central categories, as attention-constrained investors prioritize the categorical factors of larger groups.

Result 4. Trade volume is greater for larger categories

Finally, we test theoretical result 4. As in previous analyses, we perform panel regressions at both the community and asset levels. The dependent variable is the traded volume in period $t+1$, X_{mt+1}. The model specifications for both community and asset levels are as follows:

$$X_{mt+1} = \alpha + \beta N_{mt} + \gamma X_{mt} + \theta N_{posts,mt} + F_{t-1}\delta + \epsilon_{mt} \tag{45}$$

$$X_{imt+1} = \alpha + \beta N_{mt} + \theta N_{posts,imt} + F_{t-1}\delta + \lambda_i + \epsilon_{imt} \tag{46}$$

Table 3 presents the results.

Table 3: Estimation results on trade volume and community size. Assessing result 4

	Dependent variable: Log(adjusted volume)			
	Community level		Ticker level	
	(1)	(2)	(3)	(4)
Abs. community size	0.027***		0.005***	
	(0.000)		(0.000)	
Rel. community size		0.034***		0.010***
		(0.000)		(0.000)
Observations	5 492 499	5 492 499	6 377 692	6 377 692
Hour FE	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Ticker FE	No	No	Yes	Yes
R^2	0.880	0.881	0.896	0.896

Note: This table presents the estimates from equations (45) and (46). For details on the independent and control variables, please refer to Table 1. In regressions (3) and (4), we cluster standard errors at the ticker level. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table 3 supports theoretical result 4: trade volume increases with the size of the communities, reflecting X users' collective perceptions of the similarity among assets within those communities. In our theoretical model, investors allocate more attention to both categorical and asset-specific fundamentals for assets in larger categories. As a result, the *total* signal precision for these assets improves, leading to more precise investor beliefs about their payoffs. Section 3.2.4 shows that this precision channel results in higher trade volumes for those assets, as investors are more willing to bet against the market when they disagree with equilibrium prices. Therefore, the results in Table 3 align with the precision channel demonstrated in the theoretical model.

When combining the results from Table 1 and Table 3, we conclude that trade volume is higher in communities where category learning is more prevalent. This is because both theoretical result 2 of Proposition 4 and Table 3 show that categorical attention is stronger for larger categories. To assess this implication, Table 4 replaces community size with our proxies for categorical and asset-specific attention as independent variables. As expected, the results confirm that category-level attention (as well as asset-specific attention) positively impacts trade volume, hence driving the positive relationship between trade volume and community size.

Table 4: Estimation results on trade volume and attention proxies

	Dependent variable: Log(adjusted volume)					
	Community level			Ticker level		
	(1)	(2)	(3)	(4)	(5)	(6)
Categorical attention	0.184***		0.223***	0.022***		0.023***
	(0.001)		(0.001)	(0.000)		(0.000)
Asset-specific attention		0.003***	0.003***		0.000***	0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Observations	5 534 020	5 534 020	5 534 020	6 421 258	6 421 258	6 421 258
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	Yes	Yes	Yes
R^2	0.175	0.174	0.181	0.798	0.798	0.798

Note: This table presents the estimates from equations (45) and (46), but replacing community size with attention proxies. For details on the independent and control variables, please refer to Table 1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

In line with our results concerning the role of asset-specific attention on trade, Barber and Odean (2008) argue that individual investors are net buyers of attention-grabbing stocks and thus an increase in individual investor attention results in short-run positive price pressure. In the same line, Da et al. (2011) proxy attention with Google search volume (SV) to show empirically that a positive average SV index, a measure of retail attention, predicts higher short-term stock prices. Furthermore, they document stronger attention-induced price pressure among stocks in which individual investor attention matters the most.

To sum up, the empirical evidence in Tables 1, 2, 3 and 4 strongly supports the theoretical results 1 to 4 in Proposition 4, which link category-level and asset-specific attention to category size and trade volume. The analysis in this section is crucial, as it provides valuable insights that suggest that our proposed measures effectively capture category learning behavior and asset-specif attention by generating results that align with actual economic outcomes.

6 Artificial intelligence, category learning and volume

We now examine how the relationship between category-level attention and trade volume evolves with technological change, specifically focusing on the role of artificial intelligence (AI) as a major technological shift (Farboodi and Veldkamp, 2020, Abis and Veldkamp, 2024). In the model, we interpret AI as a significant relaxation of the investor's attention endowment, κ. The model predicts that a higher κ (all else being equal) leads to higher values of both categorical and asset-specific attentions in equilibrium, γ_{cm} and γ_{vm} respectively, for category $m \in \{A, B\}$ (see Appendix E.5).

Intuitively, by relaxing the attention constraint, investors can allocate more attention to both categorical and asset-specific factors, which results in greater precision for both. Additionally, since the informational content of categorical factors is larger, the ratio γ_{cm} to γ_{vm} does not decrease (see Corollary, Appendix E.5). This outcome is driven by the mechanisms of informational and strategic complementarities explained above. As a result, we expect this to impact expected trade in equilibrium, all else being equal.

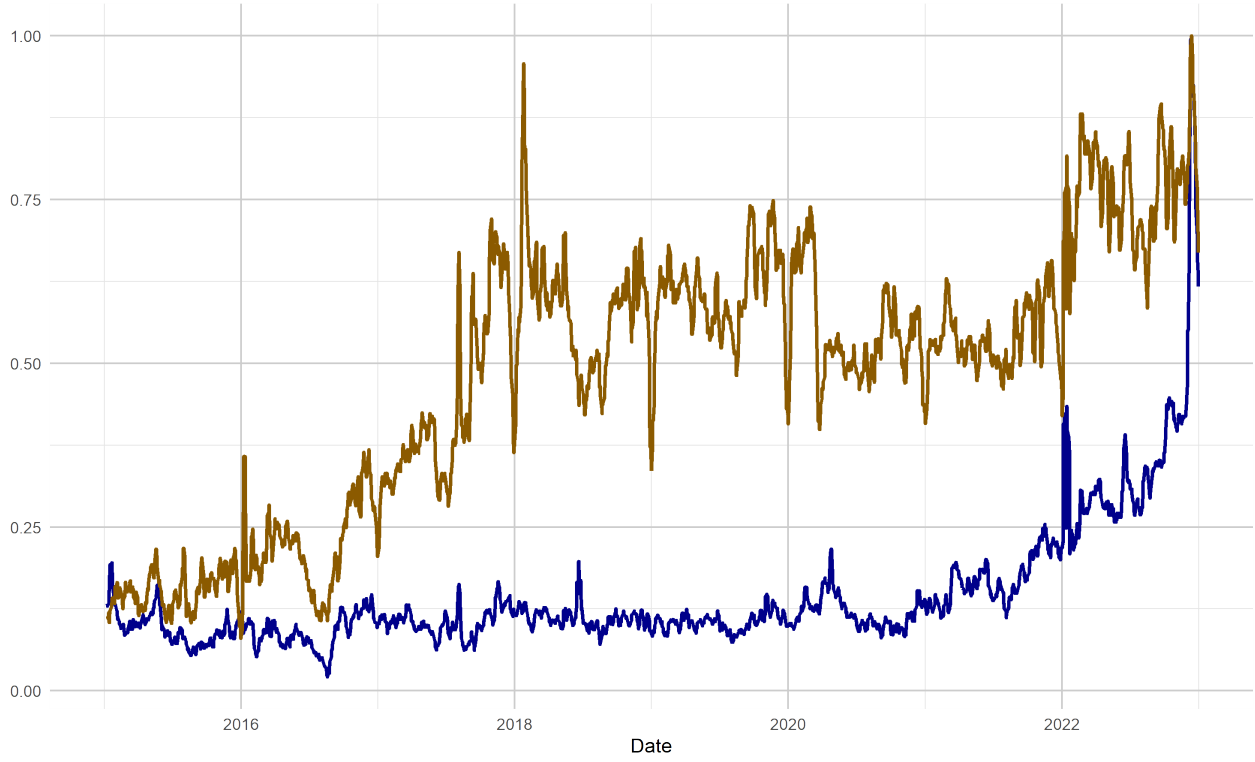
Given that the comparative statics we present in Proposition 4 hold other parameters constant, assessing the impact of AI on categorical attention and trade volume becomes an empirical question, which we address in this section. Moreover, although the model is not explicitly designed to analyze technology adoption or technological change, it offers insights to help interpret the findings that follow.

To begin with, the following figures present alternative measures of the importance of artificial intelligence. On the one hand, we use Google Trends data (http://www.google.com/trends), focusing on search frequencies for two terms: *artificial intelligence* and *AI*. We construct daily search volume indexes for both terms, which we depict in Figure 3. We compute each index by dividing the value of the search volume at a given point in time with the corresponding maximum across time. As an alternative measure of the importance of artificial intelligence, we analyze data from X. To determine the extent to which X posts from a specific period relate to a given topic (e.g., *artificial intelligence*), we follow the following steps.

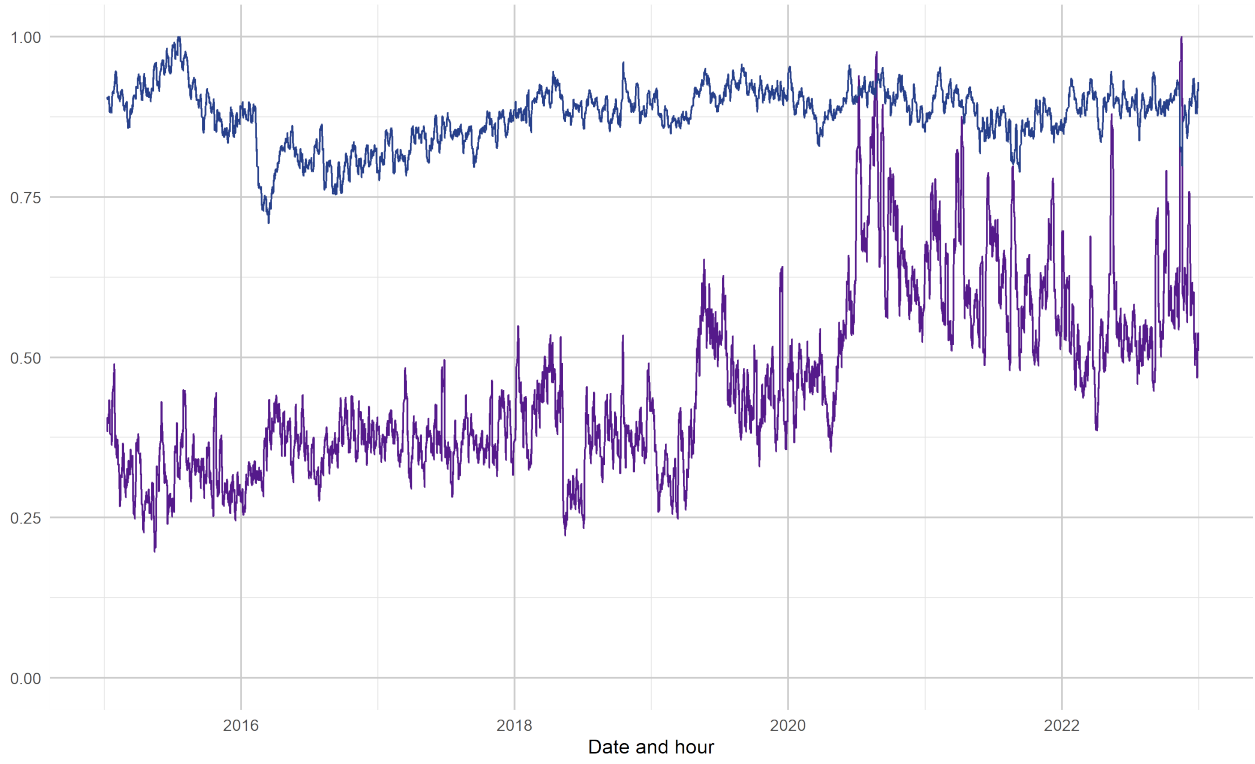
First, for each X post in the analyzed period, we calculate its semantic similarity to the target term. The terms we consider are *artificial intelligence* or *algorithmic trading*. We measure similarity using cosine similarity between the embedding of the X post and the embedding of the target concept. Embeddings are dense vector representations of text that capture semantic meaning. We generate these embeddings using Sentence Transformers (Peyton and Unnikrishnan, 2023), a framework that leverages pre-trained language models for efficient semantic similarity comparisons between short texts. Cosine similarity values range from -1 to 1, with higher values indicating greater semantic similarity. Second, after calculating the similarity for each post, we define the presence of the concept of interest as the proportion of analyzed posts

with a semantic similarity above a threshold of 0.15.[17] Figure 3 exhibits the hourly evolution of the alternative measures of the intensity of artificial intelligence relying on X data.

Figure 3: Comparison of alternative proxies for artificial intelligence



(a) Google Trend



(b) Based on posts in X

[17] In unreported analyses, we consider alternative thresholds for robustness.

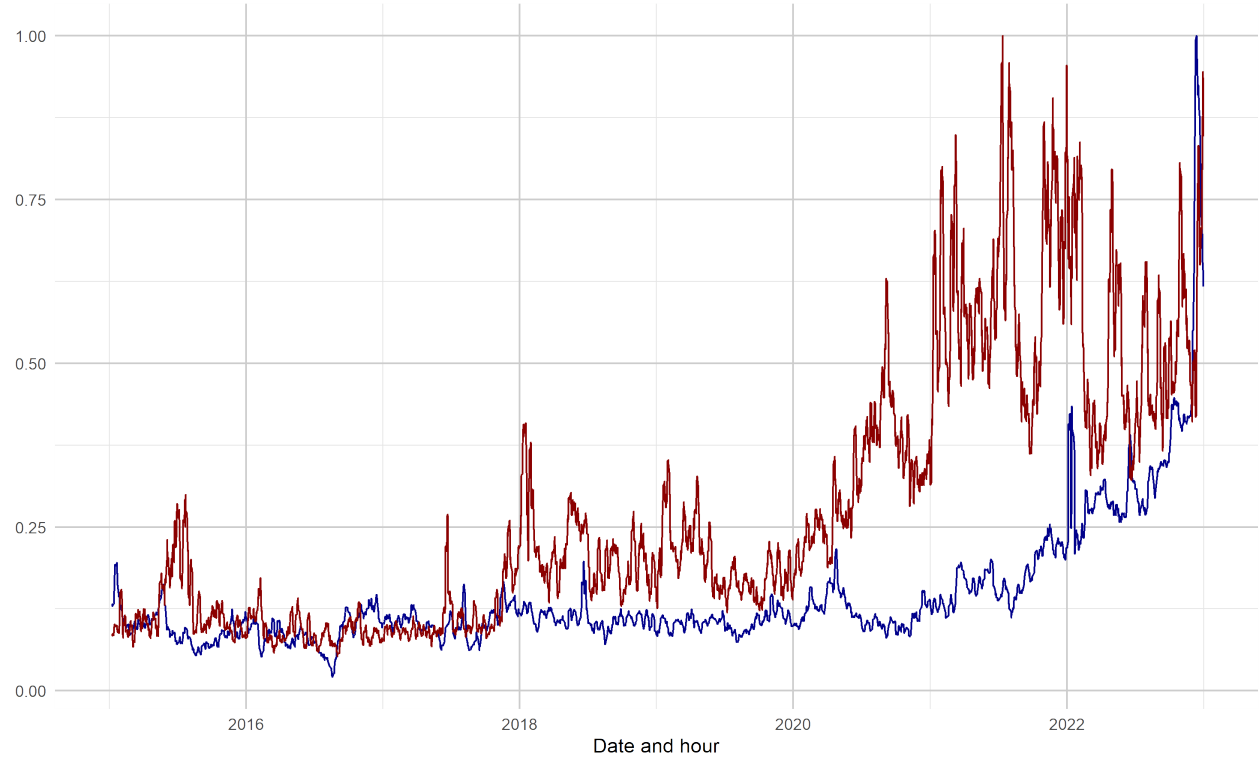
Note: Panel A of the figure depicts the evolution of alternative proxies for artificial intelligence based on Google Trends at the daily level. Each point on the curves represents the rolling average of Google search frequencies for the terms *artificial intelligence* (in brown) or *AI* (in blue). We aggregate these measures to produce rolling weekly averages for each period. To facilitate comparison among variables, we normalize the daily averages by dividing them by their respective maximum values over the sample period. Panel B presents the evolution of alternative proxies for artificial intelligence using data from X at the date-hour level. Each point on the curves represents the rolling weekly average of the proportion of X posts in a given date-hour with a semantic similarity to the concept of interest exceeding a threshold of 0.15. The terms we analyze are *artificial intelligence* (in purple) and *algorithmic trading* (in royal blue). As in Panel A, we normalize these series by dividing them by their respective maximum values over the sample period.

Figure 3 shows that, overall, the alternative measures for artificial intelligence capture a similar trend, reflecting the increasing importance of AI, with the exception of the similarity measure based on the term algorithmic trading. However, our AI proxies are subject to some caveats, which we now outline. First, a general limitation of our AI proxies, and consequently the analysis in this section, is that most do not specifically capture the application of AI technologies to financial markets. Instead, they reflect the broader rise in AI usage, as measured, for example, by search volumes for terms like artificial intelligence or AI on Google Trends. The proxy less affected by this limitation is the one based on the term algorithmic trading using similarity measures in X posts. However, as Figure 3 shows, this proxy remains relatively flat and lacks dynamic behavior over the period.

Second, the Google Trends measure for the term artificial intelligence shows an early increase that is difficult to reconcile with existing evidence on the surge of AI starting (Leitner et al., 2024). Third, the measures based on X data remain relatively flat, particularly the one that uses the term algorithmic trading. Given these observations, we prefer the Google Trends' measure that uses the keyword AI. This is the proxy we use in the benchmark regressions presented later. For robustness, we also consider the measure we construct from X posts using the term artificial intelligence.

We now begin by relating our proxy for AI intensity to category learning and, subsequently, to trade volume. To this end, Figure 4 compares the evolution of our AI proxy with category learning.

Figure 4: Artificial intelligence and categorical attention



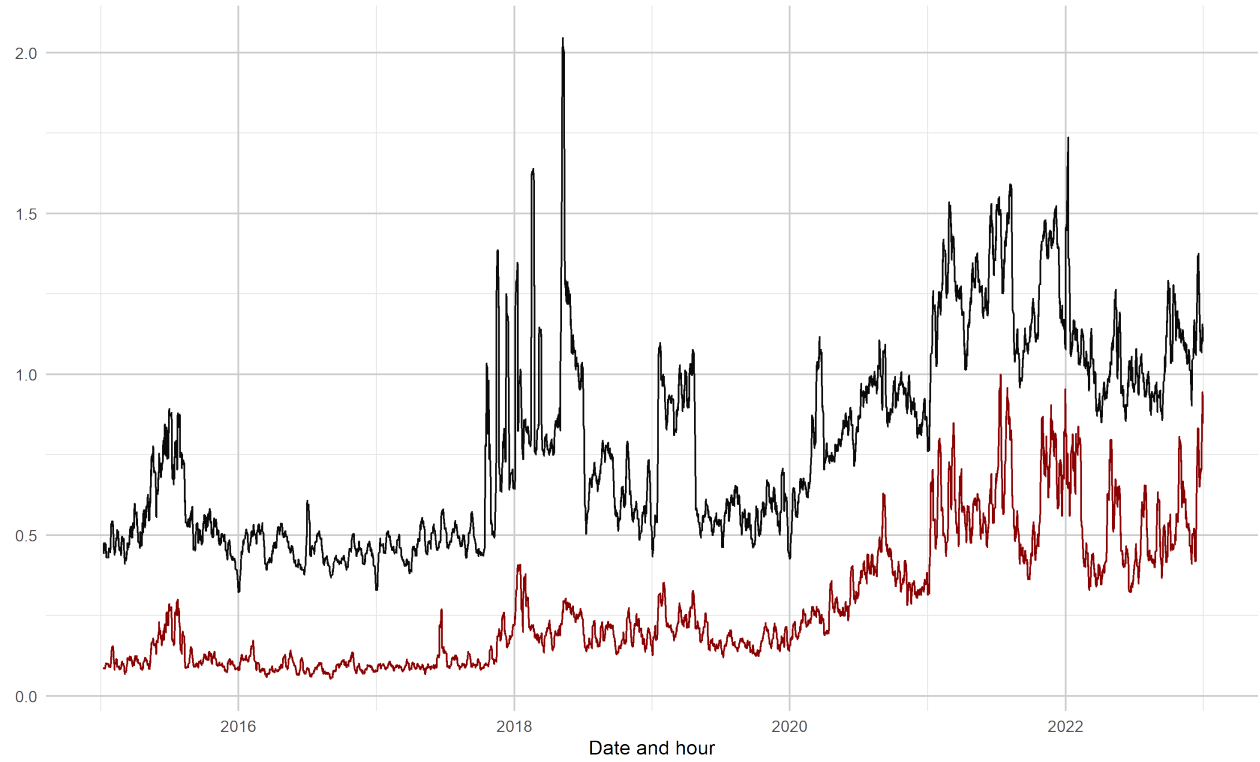
Note: The figure shows the evolution of our proxy for categorical attention (in red) at the date-hour level, alongside our preferred AI proxy (in blue). We calculate the categorical attention variable as detailed in Figure 2, with the difference that in this case, we normalize the series by dividing it by its maximum value over the sample period. For a detailed explanation of how we construct our AI variable, please refer to Figure 3.

Interestingly, Figure 4 suggests a regime shift in the evolution of categorical attention, which we associate with the rise of AI. Starting in mid-2021, we observe a noticeable increase in both the level and volatility of our proxy for categorical attention, the latter indicated by wider fluctuations. These two phenomena, which define a sort of regime shift, coincide with the significant rise of AI. Although the model is not explicitly designed to analyze technological change, it predicts that a higher κ –the theoretical counterpart of rising AI– leads to increased categorical attention. This prediction aligns with the trends we observe in Figure 4 in the later years of our study period. In the model, relaxing the attention constraint allows investors to allocate more attention to both categorical (as well as asset-specific) factors, leading to more precise beliefs about asset payoffs.

We can use our theoretical model to guide our interpretation of the rise in both the level and volatility of categorical attention following the rise in AI. Note however that in the model, the number of assets per category, N_m, remains constant. Therefore, this interpretation is rather speculative and lies beyond the boundaries of the model. A higher attention endowment (higher κ) enables investors to form more precise beliefs. However, this may also lead to greater selectivity when forming categories of similar assets. If our interpretation is correct, we should observe a reduction in the average size of communities following the rise of AI. Interestingly, Figure 5 documents a decrease in the average number of communities during 2022. Indeed,

the reduction is in the order of -45% (1/1.8-1). This suggests that rising AI not only leads to more categorical and asset-specific attention but also allows investors to become more selective (and erratic) when forming communities of perceived similar assets (those mentioned together). The combined evidence from Figures 4 and 5 is intriguing because these factors have opposing effects on trade within the model. While a larger κ leads to more category learning and more trade (assuming all else remains equal, including N_m), a reduction in community size results in less categorical attention and lower trade volume. Determining which effect dominates hence becomes an empirical question.

Figure 5: Category-level attention and community size



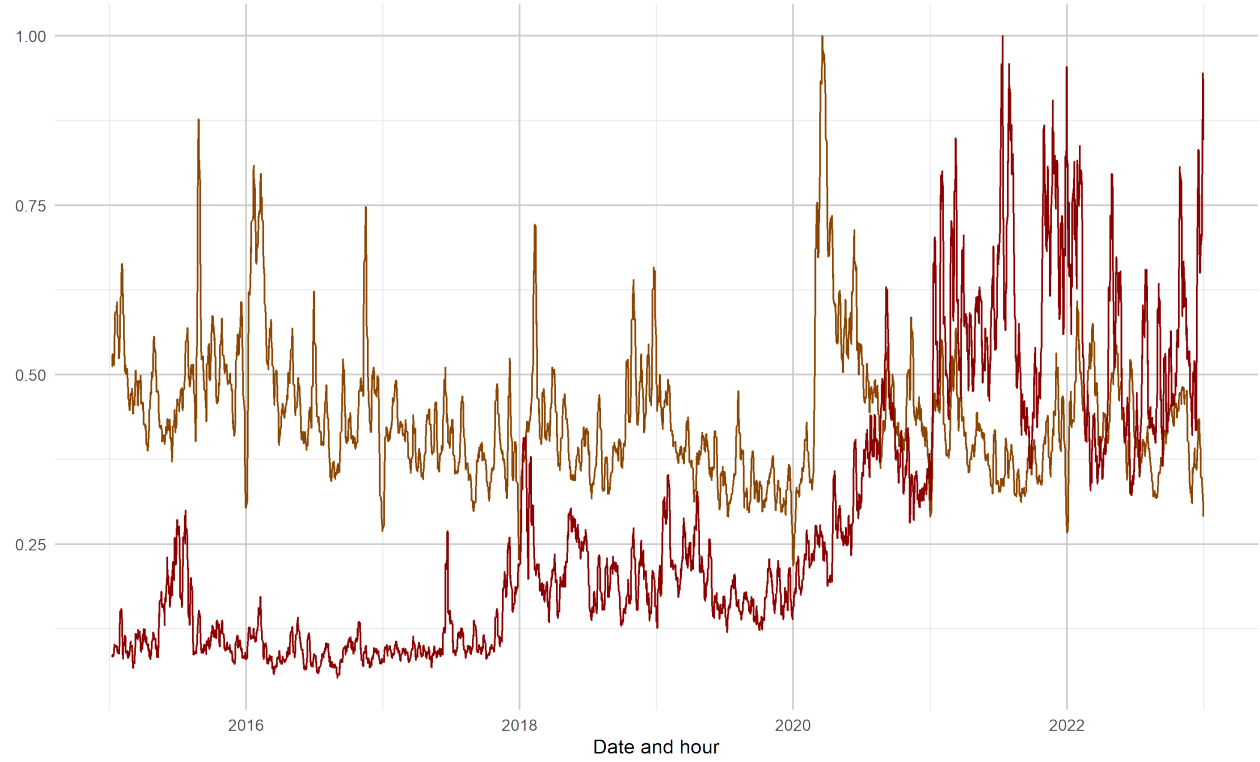
Note: The figure shows the evolution of our proxy for categorical attention (in red), alongside the average absolute community size (in black). We plot both variables at the date-hour level. For a detailed explanation of how we construct the categorical attention variable, please refer to Figure 4. We calculate absolute community size as the number of cashtags within each community and then take the average across communities (with equal weights) for every date-hour pair. We then aggregate this measure to produce a rolling weekly average for each period.

To visually inspect the relationship between categorical attention and trade volume, Figure 6 compares the evolution of our proxy for categorical attention and trade volume.[18] The figure shows that, since the rise of AI in mid-2021, trade volume has decreased, coinciding with the increase in the level and volatility of our proxy for categorical attention. While proving causality is beyond the scope of this paper, Figure 6 suggests a negative correlation between these variables for the period following 2020 (before 2020, the same correlation appears to be positive). However, it is important to acknowledge that other factors influencing trade volume

[18]For completeness, Figures A2 and A3 in the appendix exhibit the relationship between trade volume and our proxy for AI and between the standard deviation of trade volume and our proxy for AI, respectively.

may not be accounted for in this analysis. For instance, several studies documented the surge in trade volume that we observe in the data following the COVID-19 (Chiah and Zhong, 2020, Ortmann et al., 2020).[19]

Figure 6: Trade volume and category-level attention



Note: The figure depicts the evolution of our proxy for categorical attention (in red) and the average trade volume for S&P 500 assets (in dark orange). We plot both variables at the date-hour level. For details on the construction of the categorical attention variable, see Figure 4. We calculate average trading volume for each date-hour pair, using equal weights. We then aggregate this measure to produce a rolling weekly average for each period, and divide the resulting series by its maximum value over the sample period.

In summary, rising AI increases attention, enhances the precision of beliefs, and, all else equal, should lead to higher trade. However, the decrease in average community size —driven by better technologies and data enabling more selective asset categorization— counteracts this effect (Figure A4 in the appendix shows that the number of communities has not significantly changed over the period). How these two opposing forces affect trade volume remains an empirical question. We now present a set of panel regression analyses to formalize the empirical evidence provided in Figures 4, 5, and 6.

The following model specifications consider three independent variables: our proxy for categorical attention, trade volume and trade volume volatility. As in the analysis in Section 5, we perform the regressions at both the community and ticker levels, with the exception of the regressions for trade volume volatility, which are only at date-hour level. For the sake of

[19]Interestingly, Ortmann et al. (2020) document that during the COVID-19 pandemic, investors (particularly older and male individuals) significantly increased their trading volume both in the intensive and extensive margins. Investor behavior varied, with some betting on market recovery while others engaged in short selling.

brevity, the model specifications below are written at the ticker level or date-hour level, when corresponding. All the specifications include three additional independent variables: the search volume index from Google Trends, a dummy variable representing high AI importance, and an interaction term between the high AI importance dummy and community size. This interaction term allows us to empirically test the possibility of a regime shift due to the surge in AI.

When the dependent variable is our proxy for categorical attention, the model specification writes as follows:

$$\begin{aligned}\gamma_{c,mt} = \alpha + \beta_1 N_{mt} + \beta_2 AI_t + \beta_3 D_{AI,t} + \beta_4 N_{mt} \times D_{AI,t} + \\ \theta_1 N_{posts,imt} + \theta_2 N^2_{posts,imt} + F_{t-1}\delta + \lambda_i + \epsilon_{imt}\end{aligned} \tag{47}$$

where AI_t represents the search volume index for the term AI from Google Trends, while $D_{AI,t}$ is a dummy variable that equals one when the search volume index exceeds the 75th percentile of its distribution, corresponding approximately to the period beginning in September 2021. The specification in equation (47) includes an interaction effect between the dummy variable for high importance of artificial intelligence and the size of communities.

For trade volume and the standard deviation of trade volume, the model specifications at the ticker level and date hour level, respectively, are as follows:

$$\begin{aligned}X_{imt+1} = \alpha + \beta_1 N_{mt} + \beta_2 AI_t + \beta_3 D_{AI,t} + \beta_4 N_{mt} \times D_{AI,t} + \\ \theta_1 N_{posts,imt} + \theta_2 N^2_{posts,imt} + F_{t-1}\delta + \lambda_i + \epsilon_{imt}\end{aligned} \tag{48}$$

$$\begin{aligned}\sigma_{t+1} = \alpha + \beta_1 \bar{N}_t + \beta_2 AI_t + \beta_3 D_{AI,t} + \beta_4 \bar{N}_t \times D_{AI,t} + \\ \theta_1 \bar{N}_{posts,t} + \theta_2 \bar{N}^2_{posts,t} + F_{t-1}\delta + \epsilon_t\end{aligned} \tag{49}$$

where $\bar{N}_t$ and $\bar{N}_{posts,t}$ represent the average (absolute and relative) community size and the average number of X posts per community, respectively. For each date-hour, we calculate σ_{t+1} in equation (49) by first taking the logarithm of each asset's trade volume for the following date-hour and then computing the standard deviation of these log-transformed values across assets, using equal weights. Similarly, we calculate the average absolute and relative community size by averaging these metrics (with equal weights) across communities for each date-hour. To determine the average number of posts per community, we assign to each community the number of posts mentioning at least one of its constituent assets (variable $N_{posts,mt}$ from prior regressions) and then compute the average of these values across communities, again using equal weights.

Table 5: Estimation results on categorical attention, AI and community size

	Dependent variable: Categorical attention					
	Community level			Ticker level		
	(1)	(2)	(3)	(4)	(5)	(6)
AI relevance	−0.053***	0.031***	−0.030***	−2.230***	−1.364***	−2.022***
	(0.007)	(0.006)	(0.009)	(0.021)	(0.020)	(0.026)
Rel. community size		0.604***	0.584***		3.354***	3.271***
		(0.001)	(0.001)		(0.003)	(0.003)
High AI			0.023***			0.233***
			(0.002)			(0.006)
Rel. community size × High AI			0.107***			0.442***
			(0.001)			(0.004)
Observations	6 180 640	6 180 640	6 180 640	7 097 727	7 097 727	7 097 727
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	Yes	Yes	Yes
R^2	0.805	0.822	0.823	0.886	0.904	0.904

Note: This table presents the estimates from equation (47) at the community and ticker level. AI relevance corresponds to the search volume index for the term *AI* in Google Trends. High AI is a dummy variable that equals one when the search volume index exceeds the 75th percentile of its distribution, which roughly corresponds to the period beginning in September 2021. For details on the independent and control variables, please refer to Table 1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Several key findings emerge from Table 5, where our proxy for categorical attention is the dependent variable. First, the table reveals, as anticipated, that categorical attention increases with community size. Second, on average, increasing search volume related to AI is negatively associated with categorical attention. However, the dummy variable for high AI importance, which takes the value of one during the period marking the rise of AI, is positive and significant, aligning with the sign we should expect according to the model (recall that we view rising AI as an increase in the attention endowment κ). Moreover, the interaction between community size and the high AI dummy variable is also positive and significant, providing evidence for the regime shift hypothesis of our proxy for categorical attention tied to the surge in AI. Finally, the estimates at both the community and ticker levels are qualitatively consistent.

Table 6: Estimation results on trade volume, AI and community size

	Dependent variable: Log(adjusted volume)					
	Community level			Ticker level		
	(1)	(2)	(3)	(4)	(5)	(6)
AI relevance	−0.967***	−0.897***	−0.387***	−0.610***	−0.631***	−0.275***
	(0.005)	(0.005)	(0.006)	(0.002)	(0.002)	(0.003)
Rel. community size		0.428***	0.438***		0.050***	0.055***
		(0.001)	(0.001)		(0.000)	(0.000)
High AI			−0.177***			−0.127***
			(0.001)			(0.001)
Rel. community size × High AI			−0.061***			−0.008***
			(0.001)			(0.000)
Observations	5 534 020	5 534 020	5 534 020	6 421 258	6 421 258	6 421 258
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	Yes	Yes	Yes
R^2	0.176	0.251	0.253	0.800	0.801	0.802

Note: This table presents the estimates from equation (48) at the community and ticker level. For details on the independent and control variables, please refer to Table 1 and 5. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table 6 shows that, unconditionally, community size positively impacts trade volume, consistent with the model's predictions (result 4 of Proposition 4). Second, results demonstrate that the relevance of AI, whether measured continuously or through the dummy variable, negatively affects trade volume. This finding is consistent with the evidence we depict in Figure A2, in the appendix. Moreover, the interaction term between community size and the AI importance dummy variable is, negative and significant. This suggests that during periods of high AI relevance, the positive effect of community size on trade volume diminishes, likely due to investors leveraging enhanced financial technologies to become more selective when forming communities of perceived similar assets (which they reference together in their posts on X), which in turn leads to smaller communities.[20]

To sum up, there were two counterbalancing expected effects on trade volume due to rising IA. On the one hand, increasing IA tends to enhance attention, which should in turn lead to higher trade volumes. On the other hand, the relative reduction in community size –as a result of better technologies and data enabling more selective asset categorizations– would likely lead to a decrease in trade. Results in Table 6 indicate that the second effect dominates, as the coefficients for both measures for IA importance are negative and significant.

Finally, while the model does not explicitly consider the volatility of trade volume, the next table explores the effect of AI on the standard deviation of trade volume.

[20] As a robustness, Tables A1 to A3 in the appendix consider the alternative AI measure using the data on X. The results in those tables are globally consistent with the ones reported in Tables 5 to 6.

Table 7: Estimation results on standard deviation of trade volume, AI and community size

	Dependent variable:				
	Std. dev. of log(adjusted volume)				
	(1)	(2)	(3)	(4)	(5)
AI relevance	0.032^{***}	0.032^{***}	-0.055^{***}	0.031^{***}	-0.055^{***}
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Abs. community size		0.001^{*}	0.001^{**}		
		(0.000)	(0.001)		
High AI			0.041^{***}		0.041^{***}
			(0.001)		(0.001)
Abs. community size × High AI			-0.004^{***}		
Rel. community size × High AI					-0.005^{***}
					(0.001)
Rel. community size				-0.001	0.001
				(0.000)	(0.000)
Observations	14.018	14.018	14.018	14.018	14.018
Hour FE	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	No	No
R^2	0.790	0.790	0.802	0.790	0.803

Note: This table presents the estimates from equation (49) at the date-hour level. Std. dev. stands for standard deviation. For each date-hour, we construct the dependent variable by first taking the logarithm of each ticker's trade volume for the following date-hour, and then calculating the standard deviation of the resulting values across tickers, using equal weights. We calculate absolute and relative community size as the average (with equal weights) of absolute and relative community size, respectively, across communities, for each date-hour. The control variables in the regressions include the average number of posts per community (in levels and squared), the first lag of the VIX and first lag of the COR3M. For each date-hour, we construct the average number of posts per community by first assigning to each community the number of posts that mention at least one of its constituent assets, and then taking the average of these values across communities, with equal weights. We standardize the average community size variables. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table 7 shows that, overall, the relevance of AI, both measured continuously or through the dummy variable, increases the volatility of trade volume. In addition, while the impact of community size on trade volume volatility tends to be insignificant, its interaction term with the high AI dummy variable is negative but very small.

7 Conclusions and future research directions

In this paper, we develop a noisy rational expectations model to explore the allocation of investor attention between asset-specific and categorical fundamentals. Our findings show that attention-constrained investors are more likely to focus on category-level signals, especially in larger asset categories, leading to more precise beliefs and higher trade volumes. Empirical evidence, drawn from social media data, supports our model's predictions. The increasing role of AI further amplifies this behavior, allowing investors to better process categorical information while becoming more selective in forming asset communities.

The results provide valuable insights into how AI shapes investor behavior, particularly by expanding attention capacity and increasing market selectivity. AI's influence on category learning also explains shifts in trade volume and its volatility, suggesting that technological advancements are reshaping not only the speed but also the structure of financial markets.

While our findings offer a comprehensive view of attention allocation and AI's impact on trade dynamics, several avenues for future research remain. First, further exploration is needed on how AI-driven attention allocation affects market stability, especially in environments with increased algorithmic trading. The extent to which AI-induced selectivity contributes to erratic market behavior, such as flash crashes or bubbles, remains an open question.

Second, an interesting direction would involve exploring the interaction between institutional and retail investors. Our study primarily focused on retail investor attention via social media data. Future research could integrate institutional investor behavior, using proprietary data sources like Bloomberg terminals, to see how these two groups' attention allocation influences market outcomes differently.

Appendix

Internet appendix can be downloaded from here:

https://drive.google.com/drive/folders/1OqvLEEmt10nEuYhiyZkR6yMxn0czQJvq?usp=sharing

A Additional tables

Table A1: Estimation results on categorical attention, AI and community size

	Dependent variable: Categorical attention					
	Community level			Ticker level		
	(1)	(2)	(3)	(4)	(5)	(6)
AI salience	0.116***	0.172***	0.054**	3.358***	1.787***	1.003***
	(0.013)	(0.012)	(0.018)	(0.040)	(0.037)	(0.053)
Rel. community size		0.604***	0.526***		3.356***	2.926***
		(0.001)	(0.001)		(0.003)	(0.003)
High AI			0.018***			0.067***
			(0.002)			(0.006)
Rel. community size × High AI			0.222***			1.029***
			(0.001)			(0.004)
Observations	6 180 640	6 180 640	6 180 640	7 097 727	7 097 727	7 097 727
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	Yes	Yes	Yes
R^2	0.805	0.822	0.823	0.886	0.904	0.905

Note: This table presents the estimates from equation (47) at the community and ticker level. AI salience measures the presence of AI, based on similarity indexes to the term *artificial intelligence* in X. High AI is a dummy variable that equals one when the similarity index exceeds the 75th percentile of its distribution. For details on the independent and control variables, please refer to Table 1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A2: Estimation results on trade volume, AI and community size

	Dependent variable: Log(adjusted volume)					
	Community level			Ticker level		
	(1)	(2)	(3)	(4)	(5)	(6)
AI salience	−1.286***	−1.235***	−0.781***	−0.702***	−0.760***	−0.413***
	(0.010)	(0.009)	(0.013)	(0.004)	(0.004)	(0.006)
Rel. community size		0.429***	0.449***		0.050***	0.053***
		(0.001)	(0.001)		(0.000)	(0.000)
Rel. community size × High AI			−0.071***			−0.055***
			(0.002)			(0.001)
Rel. community size × High AI			−0.065***			−0.007***
			(0.001)			(0.000)
Observations	5 534 020	5 534 020	5 534 020	6 421 258	6 421 258	6 421 258
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	Yes	Yes	Yes
R^2	0.173	0.248	0.249	0.799	0.800	0.800

Note: This table presents the estimates from equation (48) at the community and ticker level. For details on the independent and control variables, please refer to Table 1 and A1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

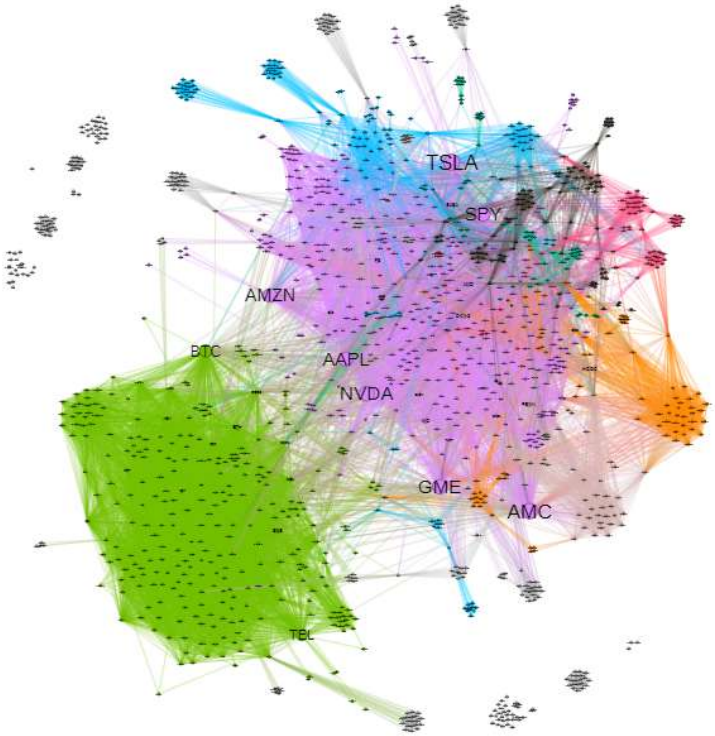
Table A3: Estimation results on standard deviation of trade volume, AI and community size

	Std. dev. across tickers, Log(adjusted volume)				
	(1)	(2)	(3)	(4)	(5)
AI salience	0.181***	0.181***	0.063***	0.181***	0.063***
	(0.008)	(0.009)	(0.012)	(0.008)	(0.012)
Abs. community size		0.000	0.000		
		(0.000)	(0.001)		
High AI			0.017***		0.018***
			(0.001)		(0.001)
Abs. community size × High AI			0.001		
			(0.001)		
Rel. community size × High AI					0.000
					(0.001)
Rel. community size				−0.001*	−0.001*
				(0.000)	(0.000)
Observations	14.018	14.018	14.018	14.018	14.018
Hour FE	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Ticker FE	No	No	No	No	No
R^2	0.796	0.796	0.799	0.796	0.799

Note: This table presents the estimates from equation (49) at the community and ticker level. For details on the independent and control variables, please refer to Tables 7 and A1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

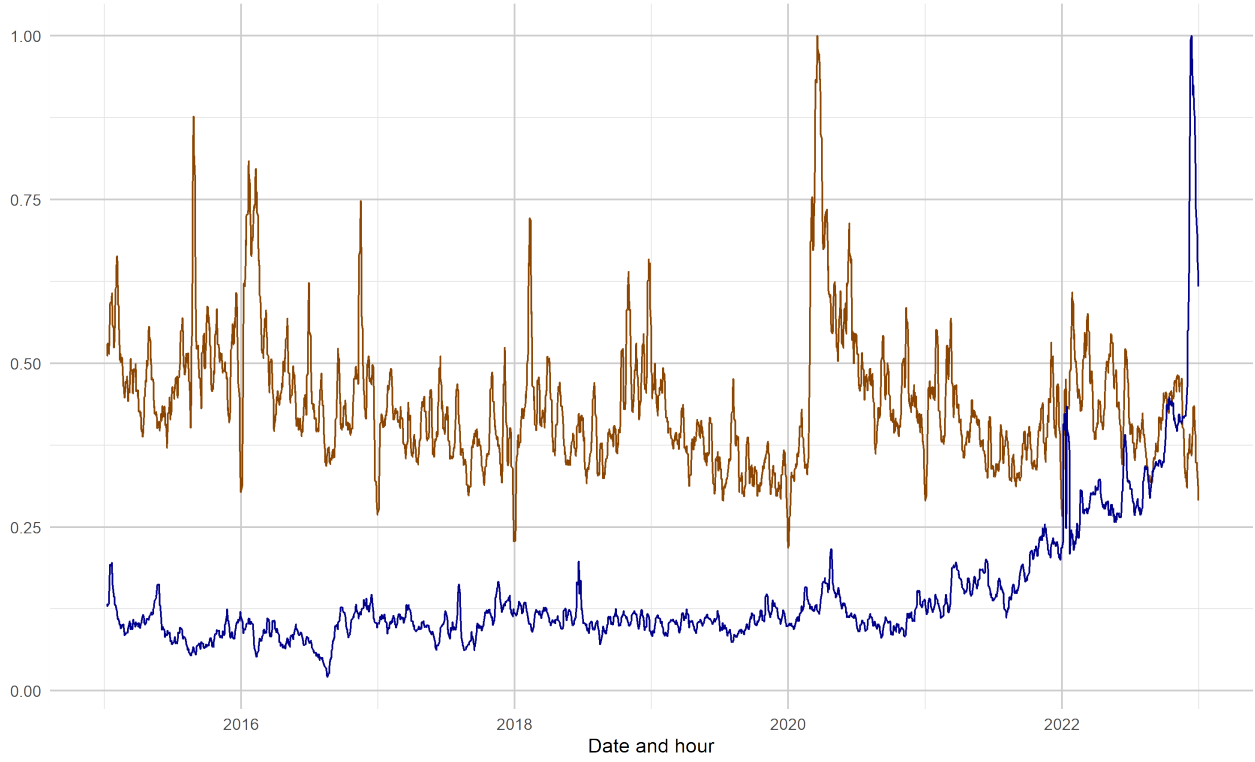
B Additional figures

Figure A1: Network of cashtags



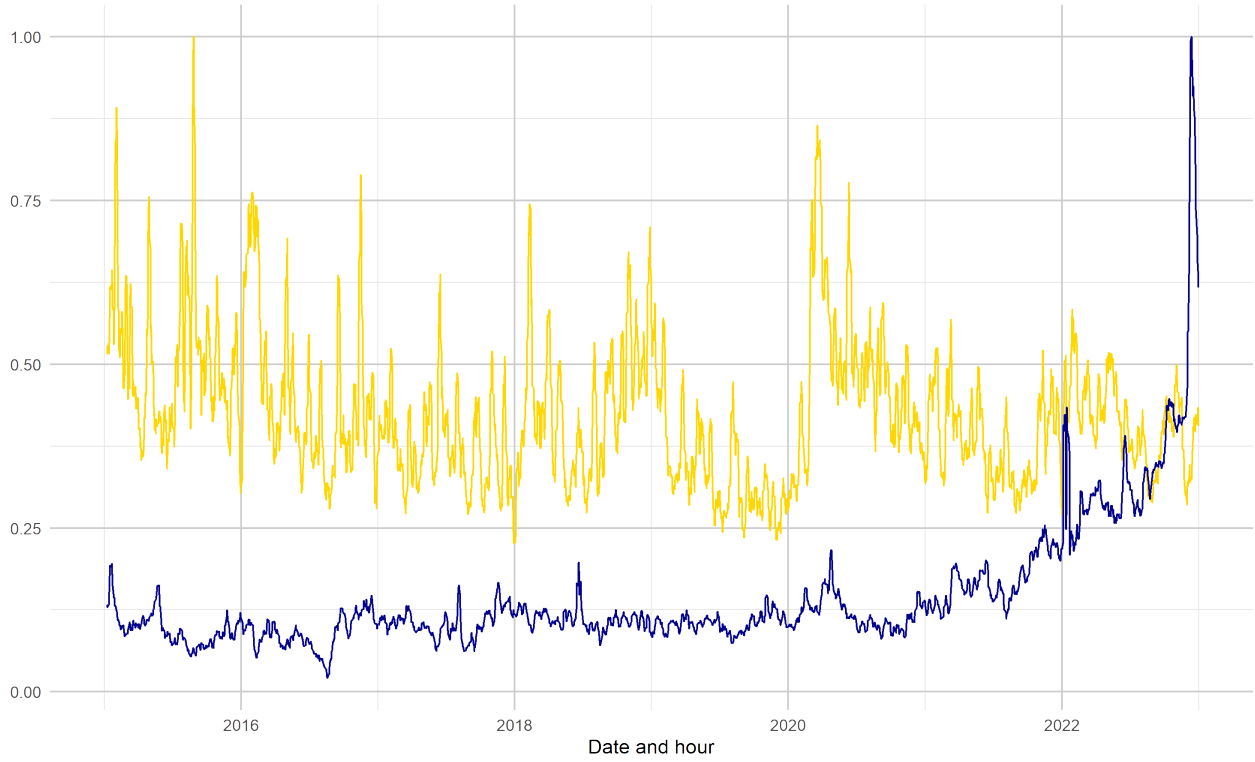
Notes: The network is created considering all X posts containing an S&P 500 ticker published from August 25, 2021, 13:00:00 EDT to August 25, 2021, 14:00:00 EDT. Node sizes are proportional to their PageRank, and colors represent communities detected using the Louvain method Blondel et al. (2008).

Figure A2: Trade volume and our AI measure



Note: The figure shows the evolution of our preferred proxy for artificial intelligence at the daily level (in blue), alongside the average trade volume for S&P 500 assets at the date-hour level (in dark orange). For details on the construction of each variable, see Figures 3 and 6, respectively.

Figure A3: Standard deviation of trade volume and our AI measure



Note: The figure illustrates the evolution of our preferred proxy for artificial intelligence at the daily level (in

blue), alongside the standard deviation of trade volume across S&P 500 assets at the date-hour level (in yellow). For details on the construction of our artificial intelligence variable, please refer to Figure 3. We calculate the standad deviation of trading volume for each date-hour pair, using equal weights. We then aggregate this measure to produce a rolling weekly average for each period, and divide the resulting series by its maximum value over the sample period.

Figure A4: Evolution of the number of communities



Note: The figure shows the evolution of the number of communities at the date hour level (in brown). We then aggregate this measure to produce a rolling monthly average for each period.

C Discussion on improper priors and learning technologies

The assumption of improper priors imposes a minimal cost on the model's ability to analyze its core focus, while significantly improving its tractability. Although proper priors could allow for richer dynamics concerning the relationship between trade and the correlation structure of priors or precision of prior beliefs, these elements are not central to our empirical application or most related theoretical literature. Moreover, this assumption does not introduce any formal deviations from standard noisy rational expectations models.

If the prior precision matrix is set to the null matrix,[21] and both the risk-aversion coefficient and risk-free asset's payoffs are set to 1, as in our model, the equilibrium price vector, the

[21] In Admati (1985), there are no priors, but rather a true distribution of asset payoffs known to all investors, with private signals referring to realizations of this distribution. This is equivalent to assuming a fixed yet unknown vector of dividends about which investors have uniform prior beliefs.

mean of each investor's beliefs, and the investor net demand function from Admati (1985) collapse to our equations (19), (20) and $\mathbf{X_1^i} = \Lambda^i[\boldsymbol{\zeta^i} - \mathbf{P_1}]$, respectively, with the latter being an intermediate step from equation (21). If we were to include uniform proper priors with mean μ_0 and precision matrix Ω_0, the only implication for the attention allocation problem would be increased complexity in the covariance matrix for equilibrium prices, complicating model resolution. Since the core relationship between signals and prices in Admati (1985) remains intact in our model, adding this complexity does not provide significant enrichment.

In the rational inattention literature following Sims (2003), it is more common to use what Van Nieuwerburgh and Veldkamp (2010) call entropy learning technologies, as in the category learning model of Peng and Xiong (2006). However, we assume additive precision learning technologies in our model for two reasons. First, Van Nieuwerburgh and Veldkamp (2010) show that entropy learning tends to result in corner solutions and, as they explain in their technical appendix, non-symmetric market equilibria where investors specialize in different assets. A model with symmetric equilibrium and interior solutions better fits our empirical application, which focuses on explaining trade patterns through more continuous shifts in attention allocation, without accounting for investor heterogeneity. Second, additive precision learning is straightforwardly compatible with the assumption of improper priors, whereas entropy learning is not. Functional forms approximating entropy learning's relationship between attention and signal precision, while retaining improper priors, would require rather peculiar assumptions.

D Expected absolute trade

Equation (22), in the case in which $\Lambda^i = \bar{\Lambda} = \Lambda$ for all $i \in [0, 1]$, can be re-written as follows:

$$\mathbf{X_1^i} = \Lambda \boldsymbol{\omega^i} + \mathbf{a} \tag{50}$$

Equation (50) implies that expected trade of any individual i is 0, since both $E(a_n) = E(\omega_n^i) = 0$ $\forall\ n \in \{1, \ldots, N\}, i \in [0, 1]$. However, this is due to selling/shorting of assets being counted as negative trade. Trade volumes of the sort analyzed in the empirical section of this paper, increase both with sales and purchases of any given asset. We are thus interested in the expected absolute value of trade in the market. To obtain this, we turn to expected trade for a given investor i.

By equation (22), trade by investor i of asset n is given by:

$$X_{n1}^i = \Lambda_n \cdot \boldsymbol{\omega^i} + a_n$$

where Λ_n represents the $n - th$ row (or, equivalently, column) vector in Λ. We are interested in $E(|X_{n1}|)$, which is given by $\sqrt{\frac{2V(X_{nt})}{\pi}}$, owing to the fact that X_{n1}^i is normally distributed.

We can write:

$$X_{n1}^i = \sum_{n'=1}^{N} \Lambda_{nn'} \omega_{n'}^i + a_n$$

which entails:

$$V(X_{n1}^i) = V\left(\sum_{m=1}^{N} \Lambda_{nn'} \omega_{n'}^i\right) + V(a_n) + 2\sum_{n'=1}^{N} \Lambda_{nn'} COV(\omega_{n'}^i a_n) = V\left(\sum_{n'=1}^{N} \Lambda_{nn'} \omega_{n'}^i\right) + \frac{1}{q_a}$$

since $COV(\omega_{n'}^i a_n) = 0 \; \forall n, n' \in \{1, \ldots, N\}$. Since $E(X_{n1}^i) = 0$, we also have that:

$$V\left(\sum_{n'=1}^{N} \Lambda_{nn'} \omega_{n'}^i\right) = E\left[\left(\sum_{n'=1}^{N} \Lambda_{nn'} \omega_{n'}^i\right)^2\right] = \sum_{n'=1}^{N} \Lambda_{nn'} \left(\sum_{k=1}^{N} \Lambda_{nk} E(\omega_{n'}^i \omega_k^i)\right) =$$
$$\sum_{n'=1}^{N} \Lambda_{nn'} \left(\sum_{k=1}^{N} \Lambda_{nk} COV(\omega_{n'}^i \omega_k^i)\right) = \sum_{n'=1}^{N} \Lambda_{nn'} \left(\Lambda_n \cdot \Lambda_{n'}^{-1}\right)$$

where $\Lambda_{n'}^{-1}$ is the m-th column of Λ^{-1}. Since Λ is the inverse of Λ^{-1}, we know that $\Lambda_n \Lambda_{n'}^{-1} = 0$ if $n \neq n'$ and $\Lambda_n \Lambda_{n'}^{-1} = 1$ if $n = n'$. Therefore, every summand in $\sum_{n'=1}^{N} \Lambda_{nn'} \left(\Lambda_n \Lambda_{n'}^{-1}\right)$ is equal to 0 except for the one in which $n' = n$, which is $\Lambda_{nn'} = \Lambda_{nn}$. We have, then, that:

$$V\left(\sum_{m=1}^{N} \Lambda_{nm} \omega_m^i\right) = \Lambda_{nn} \Rightarrow V(X_{n1}^i) = \Lambda_{nn} + \frac{1}{q_a} \Rightarrow$$
$$E(|X_{n1}^i|) = \sqrt{\frac{2\left(\Lambda_{nn} + \frac{1}{q_a}\right)}{\pi}} \tag{51}$$

Hence $E(|X_{n1}^i|)$ depends on values $2, \pi, q_a$, which are fixed throughout the model, and depends positively on Λ_{nn}, which is determined in the attention allocation stage.

It is useful to note that $\Lambda_{nn} = M_{Ann}^{-1}$ if $m(n) = A$ and $\Lambda_{nn} = M_{B,n-N_A,n-N_A}^{-1}$ if $m(n) = B$. Without loss of generality, we can consider that $\Lambda_{nn} = M_{Ann}^{-1}$. Since there is, ex-ante, no difference between assets belonging to category A, it follows that M_{Ann}^{-1} is equal for all n. Another way of looking at it is that changing the order in which assets, within a certain group i, are listed in the model, does not change M_A and hence does not change M_A^{-1}, but it does change the order of the elements in the diagonal; it follows that the diagonal of M_A^{-1} must be constant, just like that of M_A. A constant diagonal on M_A^{-1}, in turn, means that $\Lambda_{nn} = M_{Ann}^{-1} = \frac{tr\left(M_A^{-1}\right)}{N_A}$.

E Proofs

E.1 Proof of Theorem 1

E.1.1 Sketch of the proof

Using the law of iterated expectations, we write:

$$E^i\left(-e^{-W_2^i}\right) = E^i\left[E\left(-e^{-W_2^i}|\boldsymbol{\zeta^i}, \mathbf{P_1}\right)\right]$$

This allows us to separate the problem of calculating $E^i\left(-e^{-W_2^i}\right)$ into two steps. In the first one, we calculate $E^i\left(-e^{-W_2^i}|\boldsymbol{\zeta^i}, \mathbf{P_1}\right)$ given the optimal choice of $\mathbf{X_1^i}$ induced by $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$. The second step consists of allowing $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$ to vary, thus turning the solution from the first step into a random variable again, and calculating its expectation.

Using the fact that all investors hold only the risk-free asset in period $t = 1$, which yields $W_1^i = W_0^i \; \forall \; i \in [0, 1]$, we re-write equation (1) as:

$$W_2^i = W_0^i + (\mathbf{D} - \mathbf{P_1}) \cdot \mathbf{X_1^i}(\boldsymbol{\zeta^i}, \mathbf{P_1})$$

Final wealth is hence a function of the internal product of the net demand vector $\mathbf{X_1^i}\left(\boldsymbol{\zeta^i}, \mathbf{P_1}\right)$ and the vector $\mathbf{D} - \mathbf{P_1}$ of discrepancies between asset payoffs and their prices in $t = 1$. Notice that, from equation (17), we know that $\mathbf{X_1^i}\left(\boldsymbol{\zeta^i}, \mathbf{P_1}\right)$ is given by the expectation of these discrepancies, according to investor i's final beliefs, pre-multiplied by the precision of these beliefs. This yields to a situation in which both the expectation and variance of W_2^i, conditional on $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$, are quadratic forms of expected discrepancies $\mathbf{D} - \mathbf{P_1}$ according to the beliefs of investor i. Using the result that $E\left(e^Y\right) = e^{E(Y)+\frac{V(Y)}{2}}$ whenever Y is normal, we then arrive at $E^i\left(-e^{-W_2^i}|\boldsymbol{\zeta^i}, \mathbf{P_1}\right) = -e^{-u^T A u}$, for some matrix A and a vector u which depends linearly on $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$.

A result of the above is that in the second step, when $\boldsymbol{\zeta^i}$ and $\mathbf{P_1}$ are allowed to vary, u is normally distributed, hence $u^T A u$ follows a generalized chi-squared distribution; therefore, the result that $E\left(e^Y\right) = e^{E(Y)+\frac{V(Y)}{2}}$ cannot be used. We then use the moment-generating function of a generalized chi-squared distribution, together with some further algebra, to arrive at the result that $E^i\left(-e^{-W_2^i}\right) = E^i\left(-e^{-u^T A u}\right) = -\left|I_N + \frac{1}{q_a}\Lambda^i \bar{\Lambda}^{-2}\right|^{-\frac{1}{2}}$. This yields the result that maximizing $E^i\left(-e^{-W_2^i}\right)$ is equivalent to maximizing $\left|I_N + \frac{1}{q_a}\Lambda^i \bar{\Lambda}^{-2}\right|$. $\square$

E.1.2 Expectation of $-e^{-W_2^i}$ conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$

Conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$, we can use equation (1) to see that investor i's period 2 wealth is given by

$$W_2^i = W_1^i + \sum_{n=1}^{N} \left(D_n - P_{n1}\right) X_{n1}^i(\boldsymbol{\zeta^i}, \mathbf{P_1}) = W_0^i + (\mathbf{D} - \mathbf{P_1}) \cdot \mathbf{X_1^i}(\boldsymbol{\zeta^i}, \mathbf{P_1}) \Rightarrow$$
$$W_2^i = W_0^i + (\mathbf{D} - \mathbf{P_1}) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] \tag{52}$$

where we use the fact that $W_1^i = W_0^i$ because all agents hold only the risk-free asset in $t = 0$.

It is worth noting that, conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$, beliefs about $\mathbf{D}$ are described by the normal distribution previously discussed, while $\mathbf{P_1}$ is fixed, which means that beliefs about $\mathbf{D} - \mathbf{P_1}$ are also normally distributed. Since $\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)$ is also fixed conditional on $\boldsymbol{\zeta^i}, \mathbf{P_1}$, the RHS in equation (52) is a deterministic number W_0^i plus an internal product of a multivariate normal distribution and a fixed vector, and is thus normally distributed. This allows us to write:

$$E^i\left(-e^{-W_2^i} | \boldsymbol{\zeta^i}, \mathbf{P_1}\right) = -e^{-E^i\left(W_2^i | \boldsymbol{\zeta^i}, \mathbf{P_1}\right) + \frac{V^i\left(W_2^i | \boldsymbol{\zeta^i}, \mathbf{P_1}\right)}{2}}$$

In what follows it is helpful to remember that Λ^i, $\bar{\Lambda}$ and $\Omega_1^i = \Lambda^i + q_a \bar{\Lambda}^2$ are symmetric positive definite (and, as such, invertible) square matrices of dimension $N = N_A + N_B$, $\mathbf{D} - \mathbf{P_1}$ is a random vector of dimension $N_A + N_B$ with a multivariate normal distribution, and $\boldsymbol{\zeta^i} - \mathbf{P_1}$ is a fixed vector (only in this stage of the proof, when we are conditioning on ζ^i and $\mathbf{P_1}$) of dimension N.

We have that:

$$\begin{gathered}
E^i\left(W_2^i | \boldsymbol{\zeta^i}, \mathbf{P_1}\right) = E^i\left[W_0^i + (\mathbf{D} - \mathbf{P_1}) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] | \boldsymbol{\zeta^i}, \mathbf{P_1}\right] = \\
W_0^i + \left(E^i\left(\mathbf{D} | \boldsymbol{\zeta^i}, \mathbf{P_1}\right) - \mathbf{P_1}\right) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left(\left(\Omega_1^i\right)^{-1} \left[\Lambda^i \boldsymbol{\zeta^i} + q_a \bar{\Lambda}^2 \mathbf{P_1}\right] - \mathbf{P_1}\right) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left(\left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1} \left[\Lambda^i \boldsymbol{\zeta^i} + q_a \bar{\Lambda}^2 \mathbf{P_1}\right] - \mathbf{P_1}\right) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left(\left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1} \left[\Lambda^i \boldsymbol{\zeta^i} + q_a \bar{\Lambda}^2 \mathbf{P_1} - \left(\Lambda^i + q_a \bar{\Lambda}^2\right) \mathbf{P_1}\right]\right) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left(\left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1} \Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right) \cdot \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right]^T \left[\left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1}\right]^T \left[\Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)\right] = \\
W_0^i + \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)^T \left(\Lambda^i\right)^T \left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1} \Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right) = \\
W_0^i + \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)^T \Lambda^i \left(\Omega_1^i\right)^{-1} \Lambda^i \left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)
\end{gathered} \tag{53}$$

where we use the fact that $\Omega_1^i = \Lambda^i + q_a \bar{\Lambda}^2$ is a linear combination of symmetric matrices, which means that Ω_1^i is symmetric and thus so is $\left(\Omega_1^i\right)^{-1}$. We also use that Λ^i is symmetric.

In order to obtain $V^i(W_2^i | \boldsymbol{\zeta^i}, \mathbf{P_1})$, we note that $\mathbf{P_1}$ becomes fixed when conditioning on the

realization of prices and signals, and therefore $V^i(\mathbf{D}-\mathbf{P_1}|\boldsymbol{\zeta^i},\mathbf{P_1}) = V^i(\mathbf{D}|\boldsymbol{\zeta^i},\mathbf{P_1}) = (\Omega_1^i)^{-1} = (\Lambda^i+\bar{\Lambda}^2)^{-1}$. We write $V^i(\mathbf{D}-\mathbf{P_1}|\boldsymbol{\zeta^i},\mathbf{P_1}) = (\Omega_1^i)^{-1}$. Note also that, since the internal product for vector satisfies the commutative property, $(\mathbf{D}-\mathbf{P_1})\cdot\left[\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right] = \left[\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]\cdot(\mathbf{D}-\mathbf{P_1}) = \left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T(\Lambda^i)^T(\mathbf{D}-\mathbf{P_1}) = \left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i(\mathbf{D}-\mathbf{P_1})$, where the last inequality uses symmetry of Λ^i. We can thus write:

$$V^i\left(W_2^i|\boldsymbol{\zeta^i},\mathbf{P_1}\right) = V^i\left[(\mathbf{D}-\mathbf{P_1})\cdot\left[\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]|\boldsymbol{\zeta^i},\mathbf{P_1}\right] =$$
$$V^i\left(\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i(\mathbf{D}-\mathbf{P_1})|\boldsymbol{\zeta^i},\mathbf{P_1}\right) = \left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^iV^i(\mathbf{D}|\boldsymbol{\zeta^i},\mathbf{P_1})\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right) =$$
$$\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right) \tag{54}$$

From equations (53) and (54), we have that $E^i(W_2^i|\boldsymbol{\zeta^i}_1,\mathbf{P_1}) = W_0^i + V^i(W_2^i|\boldsymbol{\zeta^i}_1,\mathbf{P_1})$, which means that:

$$-E^i(W_2^i|\boldsymbol{\zeta^i}_1,\mathbf{P_1}) + \frac{V^i(W_2^i|\boldsymbol{\zeta^i}_1,\mathbf{P_1})}{2} = -W_0^i - \frac{V^i(W_2^i|\boldsymbol{\zeta^i}_1,\mathbf{P_1})}{2} =$$
$$-W_0^i - \frac{1}{2}\left[\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right] \tag{55}$$

This implies that $E^i\left(-e^{-W_2^i}|\boldsymbol{\zeta^i},\mathbf{P_1}\right) = -e^{-W_0^i-\frac{1}{2}\left[\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]}$. We now turn to $E^i\left(-e^{-W_2^i}\right) = E^i\left[E^i\left(-e^{-W_2^i}|\boldsymbol{\zeta^i},\mathbf{P_1}\right)\right] = E^i\left(-e^{-W_0^i-\frac{1}{2}\left[\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]}\right)$

E.1.3 Expectation of $-e^{-W_2^i}$ without conditioning on $\boldsymbol{\zeta^i},\mathbf{P_1}$

Since the exponent is a quadratic form of normally distributed variables, it is not itself normally distributed, which means that we can no longer use the property that $E(e^Y) = e^{E(Y)+\frac{V(Y)}{2}}$. We define:

$$\Theta^i \equiv -\frac{1}{2}\left[\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]$$

Note that $E^i\left(-e^{-W_0^i-\frac{1}{2}\left[\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i}-\mathbf{P_1}\right)\right]}\right) = E^i\left(-e^{-W_0^i+\Theta^i}\right) = E^i\left(-e^{-W_0^i}e^{\Theta^i}\right) = e^{-W_0^i}E^i\left(-e^{\Theta^i}\right)$. Since $e^{-W_0^i}$ is a deterministic positive scalar, maximizing $e^{-W_0^i}E^i\left(-e^{\Theta^i}\right)$ is equivalent to maximizing $E^i\left(-e^{\Theta^i}\right)$, or equivalently minimizing $E^i\left(e^{\Theta^i}\right)$, so we can focus purely on this last expectation.

The moment generating function of the quadratic form $\mathbf{v}^TA\mathbf{v}$, where $\mathbf{v}$ is a normally distributed vector of dimension N and with a covariance matrix equal to $\Sigma_{\mathbf{v}}$, and A is a symmetric $N\times N$ matrix, is given by:

$$M_{\mathbf{v}^TA\mathbf{v}}(s) = \frac{1}{|I_N-2sA\Sigma_{\mathbf{v}}|^{\frac{1}{2}}}e^{\frac{1}{2}E(\mathbf{v})\left[I_N-(I_N-2sA\Sigma_{\mathbf{v}})^{-1}\right]\Sigma_{\mathbf{v}}^{-1}E(\mathbf{v})}$$

If we set $A = \Lambda^i(\Omega_1^i)^{-1}\Lambda^i$, $\mathbf{v} = \boldsymbol{\zeta^i}-\mathbf{P_1} = \mathbf{D}+\boldsymbol{\omega^i}-\left(\mathbf{D}-\bar{\Lambda}\mathbf{a}\right) = \boldsymbol{\omega^i}+\bar{\Lambda}^{-1}\mathbf{a}$, and note that in

this case $E(\mathbf{v}) = 0$ and $\Sigma_{\mathbf{v}} = V(\mathbf{v}) = V\left(\boldsymbol{\omega^i} + \bar{\Lambda}^{-1}\mathbf{a}\right)$, the previous expression reduces to:

$$M_{\mathbf{v}^T A\mathbf{v}}(s) = \frac{1}{|I_N - 2sA\Sigma_{\mathbf{v}}|^{\frac{1}{2}}} \tag{56}$$

Note that, in this case, $\Sigma_{\mathbf{v}} = V\left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right) = V\left(\boldsymbol{\omega^i} + \bar{\Lambda}^{-1}\mathbf{a}\right)$. The expression in equation (56) is equal to $e^{s\mathbf{v}^T A\mathbf{v}}$, by definition of the moment generating function. Taking $s = -\frac{1}{2}$, we have that $s\mathbf{v}^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\mathbf{v} = -\frac{1}{2}\left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right)^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\left(\boldsymbol{\zeta^i} - \mathbf{P_1}\right) = \Theta$. Equation (56) then implies:

$$E^i\left[e^{\Theta}\right] = E^i\left[e^{-\frac{1}{2}\mathbf{v}^T\Lambda^i\left(\Omega_1^i\right)^{-1}\Lambda^i\mathbf{v}}\right] = \frac{1}{\left|I_N - 2\left(-\frac{1}{2}\right)A\Sigma_{\mathbf{v}}\right|^{\frac{1}{2}}} = \frac{1}{|I_N + A\Sigma_{\mathbf{v}}|^{\frac{1}{2}}} \tag{57}$$

E.1.4 Simplifying $A\Sigma_{\mathbf{v}}$

Recall that $A = \Lambda^i(\Omega_1^i)^{-1}\Lambda^i$ and $\Omega_1^i = \Lambda^i + q_a\bar{\Lambda}^2$ and therefore $A = \Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1}\Lambda^i$. In this subsection we prove that $A\Sigma_{\mathbf{v}} = \Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1}\Lambda^i\Sigma_{\mathbf{v}} = \frac{1}{q_a}\Lambda\bar{\Lambda}^{-2}$ by breaking the expression up into two factors, namely, $\Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1}$ and $\Lambda^i\Sigma_{\mathbf{v}}$, and proving separate results for each.

For $\Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1}$, we have that:

$$\left[\Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1}\right]^{-1} = \left(\Lambda^i + q_a\bar{\Lambda}^2\right)\left(\Lambda^i\right)^{-1} = I_N + q_a\bar{\Lambda}^2\left(\Lambda^i\right)^{-1}$$

which implies:

$$\Lambda^i\left(\Lambda^i + q_a\bar{\Lambda}^2\right)^{-1} = \left(I_N + q_a\bar{\Lambda}^2\left(\Lambda^i\right)^{-1}\right)^{-1} \tag{58}$$

As for $\Lambda^i\Sigma_{\mathbf{v}}$, note that:

$$\Sigma_{\mathbf{v}} = V(\mathbf{v}) = V\left(\boldsymbol{\omega^i} + \bar{\Lambda}^{-1}\mathbf{a}\right) = V\left(\boldsymbol{\omega^i}\right) + V\left(\bar{\Lambda}^{-1}\mathbf{a}\right) = \left(\Lambda^i\right)^{-1} + \bar{\Lambda}^{-1}V(\mathbf{a})\left(\bar{\Lambda}^{-1}\right)^T = \left(\Lambda^i\right)^{-1} + \bar{\Lambda}^{-1}\frac{1}{q_a}I_N\bar{\Lambda}^{-1} = \left(\Lambda^i\right)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2}$$

and thus:

$$\Lambda^i\Sigma_{\mathbf{v}} = \Lambda^i\left(\left(\Lambda^i\right)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2}\right) = I_N + \frac{1}{q_a}\Lambda^i\bar{\Lambda}^{-2}$$

Furthermore, by noting that $\left(q_a\bar{\Lambda}^2\left(\Lambda^i\right)^{-1}\right)^{-1} = \frac{1}{q_a}\Lambda^i\bar{\Lambda}^{-2}$, we obtain:

$$\Lambda^i\Sigma_{\mathbf{v}} = I_N + \frac{1}{q_a}\Lambda^i\bar{\Lambda}^{-2} = \left(q_a\bar{\Lambda}^2\left(\Lambda^i\right)^{-1} + I_N\right)\left(\frac{1}{q_a}\Lambda^i\bar{\Lambda}^{-2}\right) = \left(I_N + q_a\bar{\Lambda}^2\left(\Lambda^i\right)^{-1}\right)\left(\frac{1}{q_a}\Lambda^i\bar{\Lambda}^{-2}\right) \tag{59}$$

Finally, by combining equations (58) and (59) we obtain:

$$A\Sigma_{\mathbf{v}} = \Lambda^i \left(\Lambda^i + q_a \bar{\Lambda}^2\right)^{-1} \Lambda^i \Sigma_v =$$
$$\left(I_N + q_a \bar{\Lambda}^2 \left(\Lambda^i\right)^{-1}\right)^{-1} \left(I_N + q_a \bar{\Lambda}^2 \left(\Lambda^i\right)^{-1}\right) \left(\frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2}\right) = \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2}$$

E.1.5 Final solution for $E^i(e^{\Theta})$

With the previous result, we can re-write equation (57) as:

$$E^i\left[e^{\Theta}\right] = \frac{1}{\left|I_N + \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2}\right|^{\frac{1}{2}}} \tag{60}$$

Minimizing the RHS in equation (60) is equivalent to maximizing $\left|I_N + \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2}\right|$, which concludes the proof. □

E.2 Proof of Proposition 2

Since $(p(\bar{\gamma}_{vA}), p(\bar{\gamma}_{cA}), p(\bar{\gamma}_{vB}), p(\bar{\gamma}_{cB})) >> 0$, all elements of $\bar{\Lambda}$ are finite.

Note that $I_N + \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2} = \Lambda^i \left((\Lambda^i)^{-1} + \frac{1}{q_a} \bar{\Lambda}^{-2}\right)$. By using the properties that $|MN| = |M||N|$ for square matrices M, N of equal size and $|M^{-1}| = |M|^{-1}$, we have that:

$$\left|I_N + \frac{1}{q_a} \Lambda^i \bar{\Lambda}^{-2}\right| = |\Lambda^i| \left|(\Lambda^i)^{-1} + \frac{1}{q_a} \bar{\Lambda}^{-2}\right| = \frac{\left|(\Lambda^i)^{-1} + \frac{1}{q_a} \bar{\Lambda}^{-2}\right|}{|(\Lambda^i)^{-1}|} \tag{61}$$

Recall that:

$$\left(\Lambda^i\right)^{-1} = \begin{pmatrix} M_A^i & 0_{N_A \times N_B} \\ 0_{N_B \times N_A} & M_B^i \end{pmatrix}$$

where $0_{R \times C}$ denotes a $R \times C$ matrix filled with 0's and M_A^i, M_B^i are square matrices of size N_A and N_B, respectively. This means that the denominator on the RHS of equation (61) is given by:

$$|\left(\Lambda^i\right)^{-1}| = |M_A||M_B| \tag{62}$$

As for the numerator, we first note that $\bar{\Lambda}$ represents the precision matrix for any agent $j \in [0,1] \backslash \{i\}$, corresponding to attention vector $\bar{\boldsymbol{\gamma}}$, and as such its structure is analogous to that of Λ^i:

$$\bar{\Lambda}^{-1} = \begin{pmatrix} \bar{M}_A & 0_{N_A \times N_B} \\ 0_{N_B \times N_A} & \bar{M}_B \end{pmatrix}$$

It is easy to see that:

$$\bar{\Lambda}^{-2} = \begin{pmatrix} \bar{M}_A^2 & 0_{N_A \times N_B} \\ 0_{N_B \times N_A} & \bar{M}_B^2 \end{pmatrix}$$

and therefore:

$$(\Lambda^i)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2} = \begin{pmatrix} M_A^i + \frac{1}{q_a}\bar{M}_A^2 & 0_{N_A \times N_B} \\ 0_{N_B \times N_A} & M_B^i + \frac{1}{q_a}\bar{M}_B^2 \end{pmatrix} \Rightarrow$$
$$\left|(\Lambda^i)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2}\right| = \left|M_A^i + \frac{1}{q_a}\bar{M}_A^2\right| \left|M_B^i + \frac{1}{q_a}\bar{M}_B^2\right| \tag{63}$$

Combining equations (62) and (63), we arrive at:

$$\frac{\left|(\Lambda^i)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2}\right|}{|(\Lambda^i)^{-1}|} = \frac{\left|M_A^i + \frac{1}{q_a}\bar{M}_A^2\right| \left|M_B^i + \frac{1}{q_a}\bar{M}_B^2\right|}{|M_A^i||M_B^i|} = \left(\frac{\left|M_A^i + \frac{1}{q_a}\bar{M}_A^2\right|}{|M_A^i|}\right) \left(\frac{\left|M_B^i + \frac{1}{q_a}\bar{M}_B^2\right|}{|M_B^i|}\right) \tag{64}$$

We now turn our attention to the quotients $\frac{\left|M_m^i + \frac{1}{q_a}\bar{M}_m^2\right|}{|M_m^i|}$, for $m \in \{A, B\}$.

Recall that, for interior solutions, the symmetry between assets within the same category $m \in \{A, B\}$ means that $\gamma_{vn}^i = \gamma_{vn'}^i = \gamma_{vm}^i$ for all n, n' such that $m(n) = m(n') = m$. In this case, we have:

$$M_m^i = \begin{pmatrix} \frac{1}{p(\gamma_{vm}^i)} + \frac{1}{p(\gamma_{cm}^i)} & \frac{1}{p(\gamma_{cm}^i)} & \cdots & \frac{1}{p(\gamma_{cm}^i)} \\ \frac{1}{p(\gamma_{cm}^i)} & \frac{1}{p(\gamma_{vm}^i)} + \frac{1}{p(\gamma_{cm}^i)} & \cdots & \frac{1}{p(\gamma_{cm}^i)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p(\gamma_{cm}^i)} & \frac{1}{p(\gamma_{cm}^i)} & \cdots & \frac{1}{p(\gamma_{vm}^i)} + \frac{1}{p(\gamma_{cm}^i)} \end{pmatrix}$$

Since $\bar{\boldsymbol{\gamma}}$, the optimal solution for any agent $j \in [0,1]\backslash\{i\}$, is also assumed to be an interior solution in equilibrium, we also have $\bar{\gamma}_{vn} = \bar{\gamma}_{vn'} = \bar{\gamma}_{vm}$ for all n, n' such that $m(n) = m(n') = m$. Thus, for $m \in \{A, B\}$, the structure of $\bar{M}_m$ is analogous to that of :

$$\bar{M}_m = \begin{pmatrix} \frac{1}{p(\bar{\gamma}_{vm})} + \frac{1}{p(\bar{\gamma}_{cm})} & \frac{1}{p(\bar{\gamma}_{cm})} & \cdots & \frac{1}{p(\bar{\gamma}_{cm})} \\ \frac{1}{p(\bar{\gamma}_{cm})} & \frac{1}{p(\bar{\gamma}_{vm})} + \frac{1}{p(\bar{\gamma}_{cm})} & \cdots & \frac{1}{p(\bar{\gamma}_{cm})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p(\bar{\gamma}_{cm})} & \frac{1}{p(\bar{\gamma}_{cm})} & \cdots & \frac{1}{p(\bar{\gamma}_{vm})} + \frac{1}{p(\bar{\gamma}_{cm})} \end{pmatrix}$$

Note that, for $m \in \{A, B\}$:

$$M_m^i = \left(\frac{1}{p(\gamma_{vm}^i)}\right) I_{N_m} + \left(\frac{1}{p(\gamma_{cm}^i)}\right) Q_{N_m} \tag{65}$$

where Q_{N_m} is an $N_m \times N_m$ matrix of ones. $\bar{M}_m$ can be written the same way, except replacing γ^i_{vm} and γ^i_{vm} with $\bar{\gamma}_{vm}$ and $\bar{\gamma}_{vm}$, respectively. Therefore:

$$\begin{aligned}
\bar{M}_m^2 = \left[\left(\frac{1}{p(\bar{\gamma}_{vm})}\right) I_{N_m} + \left(\frac{1}{p(\bar{\gamma}_{cm})}\right) Q_{N_m}\right]\left[\left(\frac{1}{p(\bar{\gamma}_{vm})}\right) I_{N_m} + \left(\frac{1}{p(\bar{\gamma}_{cm})}\right) Q_{N_m}\right] = \\
\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)^2 I_{N_m} + 2\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)\left(\frac{1}{p(\bar{\gamma}_{cm})}\right) I_{N_m} Q_{N_m} + \left(\frac{1}{p(\bar{\gamma}_{cm})}\right)^2 Q_{N_m} Q_{N_m} = \\
\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)^2 I_{N_m} + 2\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)\left(\frac{1}{p(\bar{\gamma}_{cm})}\right) Q_{N_m} + N_m\left(\frac{1}{p(\bar{\gamma}_{cm})}\right)^2 Q_{N_m} = \\
\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)^2 I_{N_m} + \left[2\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)\left(\frac{1}{p(\bar{\gamma}_{cm})}\right) + N_m\left(\frac{1}{p(\bar{\gamma}_{cm})}\right)^2\right] Q_{N_m}
\end{aligned} \tag{66}$$

where the the third equality follows from the fact that, since Q_{N_m} is an $N_m \times N_m$ matrix of ones, $Q_{N_m} Q_{N_m} = N_m Q_{N_m}$.

We then have that:

$$\begin{aligned}
M^i_m + \frac{1}{q_a}\bar{M}^2_m = \\
\left(\frac{1}{p(\gamma^i_{vm})}\right) I_{N_m} + \left(\frac{1}{p(\gamma^i_{cm})}\right) Q_{N_m} + \frac{1}{q_a}\left[\left(\frac{1}{p(\bar{\gamma}_{vm})}\right)^2 I_{N_m} + \left[\frac{2}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})} + \frac{N_m}{[p(\bar{\gamma}_{cm})]^2}\right] Q_{N_m}\right] = \\
\left(\frac{1}{p(\gamma^i_{vm})} + \frac{1}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right) I_{N_m} + \left[\frac{1}{p(\gamma^i_{cm})} + \frac{2}{q_a p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})} + \frac{N_m}{q_a\left[p(\bar{\gamma}_{cm})\right]^2}\right] Q_{N_m}
\end{aligned} \tag{67}$$

Equations (65) and (67) imply that both M^i_m and $M^i_m + \frac{1}{q_a}\bar{M}^2_m$ can be written as matrices of the form:

$$M = dI_{N_m} + cQ_{N_m} \tag{68}$$

with $c, d \in \mathbb{R}$. The results we arrive at in what follows allow us to compute the determinant of any such matrix.

Lemma 1. *Let M be a $k \times k$ matrix given by $M = dI_k + cQ_k$, where $c, d \in \mathbb{R}$, $k \in \mathbb{N}$ and Q_k is a $k \times k$ matrix of ones. Then M has one eigenvalue equal to $(d + ck)$, with an associated eigenspace of dimension 1, and another eigenvalue given by d, the eigenspace of which is of dimension $k - 1$.*

Proof:

The matrix $H \equiv cQ_k$ has k equal rows given by $(c, \ldots, c)$. This means that it has rank 1, and by the rank-nullity theorem this in turn means that the kernel of H is of dimension $k - 1$. Additionally, if $\mathbb{1}_k$ is a k-sized vector of ones, then $H\mathbb{1}_k = cQ_k\mathbb{1}_k = ck\mathbb{1}_k$, hence the only non-zero eigenvalue of H is ck, and the base of its associated eigenspace is $\{\mathbb{1}_k\}$.

For any vector u within the kernel of H, we have that $Hu = 0 \Rightarrow Mu = (dI_k + H)\, u =$

$dI_k u + Hu = dI_k u = du$, which means that u is an eigenvector of M and its associated eigenvalue is d. Since the dimension of H's kernel is $k-1$, and all vectors in H's kernel belong to the eigenspace of d as an eigenvalue of M, said eigenspace is at least of dimension $k-1$.

At the same time, $M\mathbb{1}_k = (dI_k + H)\mathbb{1}_k = dI_k\mathbb{1}_k + H\mathbb{1}_k = d\mathbb{1}_k + ck\mathbb{1}_k = (d+ck)\mathbb{1}_k$. This means that $\mathbb{1}_k$ is an eigenvector of M, associated with the eigenvalue $d+ck$. The sum of the dimensions of the eigenspaces associated with d and $d+ck$ cannot exceed k, and we know that the dimension of the former is at least $k-1$ and that of the latter is at least 1. We can then conclude that the dimension of the eigenspace associated with d is $k-1$ and the dimension of the eigenspace for $d+ck$ is 1. □

Corollary 1. *For a matrix M as defined in the hypotheses of Lemma 1, we have $|M| = d^{k-1}(d+ck) = d^k + kd^{k-1}c$ and $tr(M) = dk + ck$.*

Proof:

This follows directly from Lemma 1 and the fact that the trace of a matrix equals the sum of its eigenvalues and the determinant of a matrix equals their product. □

Another consequence of Lemma 1 is the following corollary. While it is not used in this section, it proves helpful in the proof of Proposition 4, as it allows us to prove that expected absolute trade is higher for the larger group of assets in equilibrium.

Corollary 2. *For a matrix M as defined in the hypotheses of Lemma 1, if $c > 0$ and $d > 0$, then $tr(M^{-1}) = \frac{k-1}{d} + \frac{1}{d+ck}$.*

Proof:

Since $c > 0$ and $d > 0$, Lemma 1 proves that all eigenvalues of M are positive. Since M is also symmetric, this means that M is positive definite and thus diagonalizable. Therefore, there exists an invertible matrix P such that $P^{-1} = P^T$ and $M = PDP^T$, where D is a diagonal matrix, with the eigenvalues of M on its diagonal. It is also the case that $tr(M) = tr(D)$.

At the same time, $M^{-1} = \left(P^T\right)^{-1} D^{-1}P^{-1} = PD^{-1}P^T$, which means that M^{-1} is diagonalized by the same matrix P, and we also have $tr(M^{-1}) = tr(D^{-1})$. We have, then, that $tr(M^{-1})$ is equal to the sum of the inverses of M's eigenvalues, that is, $\frac{1}{d}$ summed $k-1$ times plus $\frac{1}{d+ck}$. □

As mentioned, we can write M_m^i, $m \in \{A, B\}$, as a matrix of the form described in equation (68), if we take $c = \frac{1}{p(\gamma_{vm}^i)}$ and $d = \frac{1}{p(\gamma_{cm})}$. Therefore, if we also take $k = N_m$, we have that M_m^i satisfies the hypotheses in Lemma 1, and by Corollary 1 we have:

$$|M_m^i| = \left(\frac{1}{p(\gamma_{vm}^i)}\right)^{N_m - 1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}\right] \tag{69}$$

Analogously, we can apply Corollary 1 to $M_m^i + \frac{1}{q_a}\bar{M}_m$, $m \in \{A, B\}$, and obtain:

$$\left|M_m^i + \frac{1}{q_a}\bar{M}_m^2\right| =$$

$$\left(\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a\left[p(\bar{\gamma}_{vm})\right]^2} + \frac{N_m}{p(\gamma_{cm}^i)} + \frac{2N_m}{q_a p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})} + \frac{N_m^2}{q_a\left[p(\bar{\gamma}_{cm})\right]^2}\right] =$$

$$\left(\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)} + \frac{1}{q_a}\left(\frac{1}{\left[p(\bar{\gamma}_{vm})\right]^2} + \frac{2N_m}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})} + \frac{N_m^2}{\left[p(\bar{\gamma}_{cm})\right]^2}\right)\right] =$$

$$\left(\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)} + \frac{1}{q_a}\left(\frac{1}{p(\bar{\gamma}_{vm})} + \frac{N_m}{p(\bar{\gamma}_{cm})}\right)^2\right] \tag{70}$$

Combining equations (69) and (70), we have that:

$$\frac{\left|M_m^i + \frac{1}{q_a}\bar{M}_m^2\right|}{|M_m^i|} = \frac{\left(\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a[p(\bar{\gamma}_{vm})]^2}\right)^{N_m-1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)} + \frac{1}{q_a}\left(\frac{1}{p(\bar{\gamma}_{vm})} + \frac{N_m}{p(\bar{\gamma}_{cm})}\right)^2\right]}{\left(\frac{1}{p(\gamma_{vm}^i)}\right)^{N_m-1}\left[\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}\right]} =$$

$$\left(\frac{\left(\frac{1}{p(\gamma_{vm}^i)} + \frac{1}{q_a[p(\bar{\gamma}_{vm})]^2}\right)}{\frac{1}{p(\gamma_{vm}^i)}}\right)^{N_m-1}\left(\frac{\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)} + \frac{1}{q_a}\left(\frac{1}{p(\bar{\gamma}_{vm})} + \frac{N_m}{p(\bar{\gamma}_{cm})}\right)^2}{\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}}\right) =$$

$$\left(1 + \frac{p(\gamma_{vm}^i)}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left(1 + \frac{\frac{1}{q_a}\left(\frac{1}{p(\bar{\gamma}_{vm})} + \frac{N_m}{p(\bar{\gamma}_{cm})}\right)^2}{\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}}\right) =$$

$$\left(1 + \frac{p(\gamma_{vm}^i)}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left(1 + \frac{\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2}{\frac{p(\gamma_{cm}^i)+N_m p(\gamma_{vm}^i)}{p(\gamma_{vm}^i)p(\gamma_{cm}^i)}}\right) =$$

$$\left(1 + \frac{p(\gamma_{vm}^i)}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right)^{N_m-1}\left(1 + \left(\frac{p(\gamma_{vm}^i)p(\gamma_{cm}^i)}{p(\gamma_{cm}^i) + N_m p(\gamma_{vm}^i)}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm}) + N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2\right) \tag{71}$$

which, when combined with equation (64), yields:

$$\frac{\left|(\Lambda^i)^{-1} + \frac{1}{q_a}\bar{\Lambda}^{-2}\right|}{|(\Lambda^i)^{-1}|} = \left(1 + \frac{p(\gamma_{vA}^i)}{q_a\left[p(\bar{\gamma}_{vA})\right]^2}\right)^{N_A-1}\left(1 + \left(\frac{p(\gamma_{vA}^i)p(\gamma_{cA}^i)}{p(\gamma_{cA}^i) + N_A p(\gamma_{vA}^i)}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA}) + N_A p(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2\right)$$
$$\left(1 + \frac{p(\gamma_{vB}^i)}{q_a\left[p(\bar{\gamma}_{vB})\right]^2}\right)^{N_B-1}\left(1 + \left(\frac{p(\gamma_{vB}^i)p(\gamma_{cB}^i)}{p(\gamma_{cB}^i) + N_B p(\gamma_{vB}^i)}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB}) + N_B p(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2\right) \quad \square$$

E.3 Proof of Proposition 3

Uniqueness of optimal attention allocation

In this subsection we begin by proving, using the fact that $p(.)$ is concave, Lemma 2, which states that $f_{\bar{\gamma}}(\boldsymbol{\gamma^i})$ is strictly concave over its domain $\mathbb{R}^4_+$.

Lemma 2. *Given any value of $\bar{\gamma} \in \mathbb{R}^4_+$ such that $(p(\bar{\gamma}_{vA}), p(\bar{\gamma}_{cA}), p(\bar{\gamma}_{vB}), p(\bar{\gamma}_{cB})) \in \mathbb{R}^4_{++}$, investor i's objective function $f_{\bar{\gamma}}(\boldsymbol{\gamma^i})$ is strictly concave for all $\boldsymbol{\gamma^i} \in \mathbb{R}^4_+$.*

Proof:

Notice that $f_{\bar{\gamma}}(\boldsymbol{\gamma^i})$ in equation (26) can be written as

$$f_{\bar{\gamma}}(\boldsymbol{\gamma^i}) = \phi_1\left(\gamma^i_{vA}\right) + \phi_2\left(\gamma^i_{vA}, \gamma^i_{cA}\right) + \phi_3\left(\gamma^i_{vB}\right) + \phi_4\left(\gamma^i_{vB}, \gamma^i_{cB}\right)$$

where

$$\phi_1\left(\gamma^i_{vA}\right) = (N_A - 1)\, ln\left(1 + \frac{p(\gamma^i_{vA})}{q_a\left[p(\bar{\gamma}_{vA})\right]^2}\right)$$

$$\phi_2\left(\gamma^i_{vA}, \gamma^i_{cA}\right) = ln\left(1 + \left(\frac{p(\gamma^i_{vA})p(\gamma^i_{cA})}{p(\gamma^i_{cA}) + N_A p(\gamma^i_{vA})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cA}) + N_A p(\bar{\gamma}_{vA})}{p(\bar{\gamma}_{vA})p(\bar{\gamma}_{cA})}\right)^2\right)$$

$$\phi_3\left(\gamma^i_{vB}\right) = (N_B - 1)\, ln\left(1 + \frac{p(\gamma^i_{vB})}{q_a\left[p(\bar{\gamma}_{vB})\right]^2}\right)$$

$$\phi_4\left(\gamma^i_{vB}, \gamma^i_{cB}\right) = ln\left(1 + \left(\frac{p(\gamma^i_{vB})p(\gamma^i_{cB})}{p(\gamma^i_{cB}) + N_B p(\gamma^i_{vB})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cB}) + N_B p(\bar{\gamma}_{vB})}{p(\bar{\gamma}_{vB})p(\bar{\gamma}_{cB})}\right)^2\right)$$

and $\phi_1(.)$ and $\phi_3(.)$ are defined over $\mathbb{R}_+$, while $\phi_2(.)$ and $\phi_4(.)$ are defined over $\mathbb{R}^2_+$.

We firstly note that $\phi_1\left(\gamma^i_{vA}\right)$ is strictly concave:

$$\phi'(\gamma^i_{vA}) = (N_A - 1)\frac{\frac{p'(\gamma^i_{vA})}{q_a[p(\bar{\gamma}_{vA})]^2}}{1 + \frac{p(\gamma^i_{vA})}{q_a[p(\bar{\gamma}_{vA})]^2}} = \frac{(N_A - 1)p'(\gamma^i_{vA})}{q_a\left[p(\bar{\gamma}_{vA})\right]^2 + p(\gamma^i_{vA})} > 0 \Rightarrow$$

$$\phi''(\gamma^i_{vA}) = (N_A - 1)\frac{p''(\gamma^i_{vA})p(\gamma^i_{vA}) - \left[p'(\gamma^i_{vA})\right]^2}{\left[q_a\left[p(\bar{\gamma}_{vA})\right]^2 + p(\gamma^i_{vA})\right]^2} < 0$$

since $p''(\gamma^i_{vA}) \leq 0$, $p(\gamma^i_{vA}) \geq 0$ and $p'(\gamma^i_{vA}) > 0$ for all $\gamma^i_{vA} \in \mathbb{R}_+$. An analogous argument proves that $\phi_3\left(\gamma^i_{vB}\right)$ is also strictly concave.

We now prove that $\phi_2\left(\gamma^i_{vA}, \gamma^i_{cA}\right)$ and $\phi_4\left(\gamma^i_{vB}, \gamma^i_{cB}\right)$ are also concave, and strictly concave in γ_{cA} and γ_{cB}, respectively, whenever γ_{vA} and γ_{vB} (respectively) are left fixed. It is convenient to write these results as two sublemmas.

Sublemma 2.1. *Functions $\phi_2(.)$ and $\phi_4(.)$ are concave.*

Proof:

Functions $\phi_2(.)$ and $\phi_4(.)$ can be written $\phi_{k(m)}(.)$, with $k(m) = 2$ if $m = A$ and $k(m) = 4$ if $m = B$. This proof will thus be written for a generic $m \in \{A, B\}$ and its corresponding function $\phi_{k(m)}(.)$.

Firstly, note that

$$\frac{p(\gamma_{vm}^i)p(\gamma_{cm}^i)}{p(\gamma_{cm}^i) + N_m p(\gamma_{vm}^i)} = \frac{1}{\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}} = \left(\frac{1}{1+N_m}\right)\left(\frac{1+N_m}{\frac{1}{p(\gamma_{vm}^i)} + \frac{N_m}{p(\gamma_{cm}^i)}}\right) = \frac{1}{1+N_m} H_m\left(p(\gamma_{vm}^i), p(\gamma_{cm}^i)\right)$$

where $H_m\left(p(\gamma_{vm}^i), p(\gamma_{cm}^i)\right)$ is the weighted harmonic mean of $p(\gamma_{vm}^i)$ and $p(\gamma_{vm})$, with their respective weights given by 1 and N_m.

Since the weighted harmonic mean is a concave function which is increasing in each argument, we know that $H_m(x, y)$ is increasing in both arguments, as well as concave. We now prove that $H_m\left(p(\gamma_{vm}^i), p(\gamma_{cm}^i)\right)$ is also concave in γ_{vm}^i and γ_{cm}^i. Let us take two points in $\mathbb{R}_+^2$, given by $\boldsymbol{\gamma_m} = (\gamma_{vm}, \gamma_{cm})$ and $\boldsymbol{\gamma'_m} = (\gamma'_{vm}, \gamma'_{cm})$, and consider $\boldsymbol{\gamma''_m} = (\gamma''_{vm}, \gamma''_{cm}) = t\boldsymbol{\gamma_m} + (1-t)\boldsymbol{\gamma'_m}$, for some $t \in [0, 1]$. We have that:

$$H_m\left(p(\gamma''_{vm}), p(\gamma''_{cm})\right) \geq H_m\left(tp(\gamma_{vm}) + (1-t)p(\gamma'_{vm}), tp(\gamma_{cm}) + (1-t)p(\gamma'_{cm})\right) \geq$$
$$tH_m\left(p(\gamma_{vm}), p(\gamma_{cm})\right) + (1-t)H_m\left(p(\gamma'_{vm}), p(\gamma'_{cm})\right)$$

where the first inequality is due to the fact that $p(.)$ is concave and $H_m(.)$ is increasing in $p(\gamma_{vm}^i)$ and $p(\gamma_{vm}^i)$, and the second inequality is due to the fact that $H_m(.)$ is concave in $p(\gamma_{vm}^i)$ and $p(\gamma_{vm}^i)$. Since the previous result holds for any $t \in [0, 1]$ and any $\boldsymbol{\gamma_m}, \boldsymbol{\gamma'_m} \in \mathbb{R}_+^2$, this proves the concavity of $H_m(.)$.

We have, then, that $\phi_{k(m)}(.) = \left(\psi_m \circ H_m\right)(.)$, where $\psi_m : \mathbb{R} \to \mathbb{R}$ is defined as:

$$\psi_m(x) = ln\left(1 + \frac{x}{q_a(1+N_m)}\left(\frac{p(\bar{\gamma}_{cm}) + N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2\right)$$

Since $\psi_m(.)$ is concave and increasing (as it is the composition of a linear function and the natural logarithm), we have that both $\psi_m(.)$ and $H_m(.)$ are concave and $\psi_m(.)$ is increasing, hence $\phi_{k(m)}(.) = \left(\psi_m \circ H_m\right)(.)$ is concave. □

Sublemma 2.2. *Let Φ_A be the family of functions $\tilde{\phi}_{\gamma_{vA}^i} : \mathbb{R}_+ \to \mathbb{R}$ given by $\tilde{\phi}_{\gamma_{vA}^i}(\gamma_{cA}) = \phi_4\left(\gamma_{vA}^i, \gamma_{cA}^i\right)$ for a fixed $\gamma_{vA}^i \geq 0$, and Φ_B be the family of functions $\tilde{\phi}_{\gamma_{vB}^i} : \mathbb{R}_+ \to \mathbb{R}$ given by $\tilde{\phi}_{\gamma_{vB}^i}(\gamma_{cB}) = \phi_4\left(\gamma_{vB}^i, \gamma_{cB}^i\right)$ for a fixed $\gamma_{vB}^i \geq 0$. Then any function in $\Phi_A \cup \Phi_B$ is strictly concave.*

Proof:

As in the previous proof, we show this result for any function in Φ_m, for any $m \in \{A, B\}$, and write $\phi_2(.)$ and $\phi_4(.)$ as $\phi_{k(m)}$.

We first consider the family of functions $\tilde{H}_{\gamma_{vm}^i} : \mathbb{R}_+ \to \mathbb{R}$, defined as $\tilde{H}_{\gamma_{vm}^i}(\gamma_{cm}^i) = H_m\left(p(\gamma_{vm}^i), p(\gamma_{cm}^i)\right)$ for some fixed value of γ_{vm}^i, and prove that any function within this family is strictly increasing

and strictly concave. We do this by first proving the strict increasingness and strict concavity of the family of functions $\hat{H}_{xm} : \mathbb{R}_+ \to \mathbb{R}$, where $\hat{H}_{xm}(y)$ is the weighted harmonic average of y and some fixed x (with y weighted by N_m and x by 1), and then using the fact that $\tilde{H}_{\gamma^i_{vm}}(.) = \left(\hat{H}_{p(\gamma^i_{vm})m} \circ p\right)(.)$.

We have that, for any $x \in \mathbb{R}$:

$$\hat{H}_{xm}(y) = \left(\frac{1+N_m}{\frac{1}{x}+\frac{N_m}{y}}\right) = (1+N_m)\left(\frac{xy}{y+N_m x}\right)$$

Hence, for any $x \geq 0$ and over its entire domain $\mathbb{R}_+$, $\hat{H}_{xm}(.)$ is strictly increasing:

$$\hat{H}'_{xm}(y) = (1+N_m)\left(\frac{x(y+N_m x)-xy}{(y+N_m x)^2}\right) = (1+N_m)\left(\frac{N_m x^2}{(y+N_m x)^2}\right) > 0$$

and strictly concave:

$$\hat{H}''_{xm}(y) = -2(1+N_m)\left(\frac{N_m x^2}{(y+N_m x)^4}\right) < 0$$

We thus have that $\tilde{H}_{\gamma^i_{vm}}(.)$ is equal to the composition $\left(\hat{H}_{p(\gamma^i_{vm})m} \circ p\right)(.)$, $\hat{H}_{p(\gamma^i_{vm})m}(.)$ and $p(.)$ are both strictly increasing, $\hat{H}_{p(\gamma^i_{vm})m}(.)$ is strictly concave and $p(.)$ is concave. Given $\gamma_{cm}, \gamma'_{cm} \geq 0$, $\gamma_{cm} \neq \gamma'_{cm}$, and $\gamma''_{cm} = t\gamma_{cm} + (1-t)\gamma'_{cm}$ for some $t \in (0,1)$, we have:

$$\tilde{H}_{\gamma^i_{vm}}(\gamma''_{cm}) = \left(\hat{H}_{p(\gamma^i_{vm})m} \circ p\right)(\gamma''_{cm}) = \hat{H}_{p(\gamma^i_{vm})m}\left(p(\gamma''_{cm})\right) \geq \hat{H}_{p(\gamma^i_{vm})m}\left(tp(\gamma_{cm}) + (1-t)p(\gamma'_{cm})\right) > t\hat{H}_{p(\gamma^i_{vm})m}\left(p(\gamma_{cm})\right) + (1-t)\hat{H}_{p(\gamma^i_{vm})m}\left(p(\gamma'_{cm})\right)$$

where the first inequality follows from the fact that $\hat{H}_{p(\gamma^i_{vm})m}(.)$ is increasing and $p(.)$ is concave, while the second inequality follows from the combination of the strict increasingness of $\hat{H}_{p(\gamma^i_{vm})m}(.)$ and the fact that $p(.)$ is injective (since it is strictly increasing) and therefore $p(\gamma_{cm}) \neq p(\gamma'_{cm})$. Since this is true for any $t \in (0,1)$, any $\gamma_{cm}, \gamma'_{cm} \geq 0$ with $\gamma_{cm} \neq \gamma'_{cm}$ and any value of γ^i_{vm}, we have that all functions $\tilde{H}_{\gamma^i_{vm}}(.)$, for some γ^i_{vm}, are strictly concave. Additionally, for any $\gamma^i_{vm} \geq 0$ we have that $\hat{H}_{p(\gamma^i_{vm})m}(.)$ and $p(.)$ are both strictly increasing functions, so their composition $\tilde{H}_{\gamma^i_{vm}}(.)$ is also strictly increasing .

Taking again the function $\psi_m(.)$ from the proof of the last lemma, and noting that it is strictly increasing and strictly concave (not merely increasing and concave), we have that, for any given $\gamma^i_{vm} \geq 0$, $\tilde{\phi}_{\gamma^i_{vm}}(.) = \left(\psi_m \circ \tilde{H}_{\gamma^i_{vm}}\right)(.)$. Thus, for any given $\gamma_{cm}, \gamma'_{cm} \geq 0$, $\gamma_{cm} \neq \gamma'_{cm}$, and $\gamma''_{cm} = t\gamma_{cm} + (1-t)\gamma'_{cm}$ for some $t \in (0,1)$, we have that:

$$\tilde{\phi}_{\gamma^i_{vm}}(\gamma''_{cm}) = \left(\psi_m \circ \tilde{H}_{\gamma^i_{vm}}\right)(\gamma''_{cm}) = \psi_m\left(\tilde{H}_{\gamma^i_{vm}}(\gamma''_{cm})\right) > \psi_m\left(t\tilde{H}_{\gamma^i_{vm}}(\gamma_{cm}) + (1-t)\tilde{H}_{\gamma^i_{vm}}(\gamma'_{cm})\right) > t\psi\left(\tilde{H}_{\gamma^i_{vm}}(\gamma_{cm})\right) + (1-t)\left(\tilde{H}_{\gamma^i_{vm}}(\gamma'_{cm})\right)$$

where the first inequality results from the fact that $\psi(.)$ is strictly increasing and $\tilde{H}_{\gamma^i_{vm}}(.)$ is strictly concave, and the second inequality results from the fact that $\psi(.)$ is strictly increasing and $\tilde{H}_{\gamma^i_{vm}}(.)$ is strictly increasing and thus injective, which implies $\tilde{H}_{\gamma^i_{vm}}(\gamma_{cm}) \neq \tilde{H}_{\gamma^i_{vm}}(\gamma'_{cm})$. Since the previous inequality holds for any $t \in (0,1)$ and $\gamma_{vm} \geq 0$, we have that all functions in Φ_m are strictly concave. □

We now take $\boldsymbol{\gamma}, \boldsymbol{\gamma'} \in \mathbb{R}^4_+$, $\boldsymbol{\gamma} \neq \boldsymbol{\gamma'}$, written as $\boldsymbol{\gamma} = (\gamma_{vA}, \gamma_{cA}, \gamma_{vB}, \gamma_{cB})^T$ and $\boldsymbol{\gamma'} = (\gamma'_{vA}, \gamma'_{cA}, \gamma'_{vB}, \gamma'_{cB})^T$. We also take $\boldsymbol{\gamma''} = t\boldsymbol{\gamma} + (1-t)\boldsymbol{\gamma'}$ for some $t \in (0,1)$. Since $\phi_1(.)$ is strictly concave, we have:

$$\phi_1(\gamma''_{vA}) \leq t\phi_1(\gamma_{vA}) + (1-t)\phi_1(\gamma'_{vA}) \tag{72}$$

where the inequality is strict unless $\gamma_{vA} = \gamma'_{vA}$. At the same time, by Sublemma 2.1:

$$\phi_2(\gamma''_{vA}, \gamma''_{cA}) \leq t\phi_2(\gamma_{vA}, \gamma_{cA}) + (1-t)\phi_2(\gamma'_{vA}, \gamma'_{cA}) \tag{73}$$

In the special case in which $\gamma_{vA} = \gamma'_{vA} = \tilde{\gamma}_{vA}$, inequality (73) can be written:

$$\phi_2(\gamma''_{vA}, \gamma''_{cA}) = \phi_2(\tilde{\gamma}_{vA}, \gamma''_{cA}) = \tilde{\phi}_{\tilde{\gamma}_{vA}}(\gamma''_{cA}) \leq t\tilde{\phi}_{\tilde{\gamma}_{vA}}(\gamma''_{cA}) + (1-t)\tilde{\phi}_{\tilde{\gamma}_{vA}}(\gamma''_{cA}) =$$
$$t\phi_2(\tilde{\gamma}_{vA}, \gamma_{cA}) + (1-t)\phi_2(\tilde{\gamma}_{vA}, \gamma'_{cA}) = t\phi_2(\gamma_{vA}, \gamma_{cA}) + (1-t)\phi_2(\gamma'_{vA}, \gamma'_{cA})$$

with $\tilde{\phi}_{\tilde{\gamma}_{vA}} \in \Phi_A$. Since, by Sublemma 2.2, all functions in Φ_A are strictly concave, the previous inequality is strict unless $\gamma_{cA} = \gamma'_{cA}$.

We are thus in a situation in which one inequality out of (72) and (73) is strict, unless $\gamma_{vA} = \gamma'_{vA}$ and $\gamma_{cA} = \gamma'_{cA}$.

Applying the same reasoning for $\phi_3(.)$ and $\phi_4(.)$, we have:

$$\begin{gathered}
\phi_1(\gamma''_{vA}) \leq t\phi_1(\gamma_{vA}) + (1-t)\phi_1(\gamma'_{vA}) \\
\phi_2(\gamma''_{vA}, \gamma''_{cA}) \leq t\phi_2(\gamma_{vA}, \gamma_{cA}) + (1-t)\phi_2(\gamma'_{vA}, \gamma'_{cA}) \\
\phi_3(\gamma''_{vB}) \leq t\phi_3(\gamma_{vB}) + (1-t)\phi_3(\gamma'_{vB}) \\
\phi_4(\gamma''_{vB}, \gamma''_{cB}) \leq t\phi_4(\gamma_{vB}, \gamma_{cB}) + (1-t)\phi_4(\gamma'_{vB}, \gamma'_{cB})
\end{gathered}$$

where at least one of the inequalities is strict unless $\boldsymbol{\gamma} = \boldsymbol{\gamma'}$, which we have assumed is not the case. Therefore, adding up the four previous inequalities, and bearing in mind that one of them is strict, we have:

$$\begin{gathered}
\phi_1(\gamma''_{vA}) + \phi_2(\gamma''_{vA}, \gamma''_{cA}) + \phi_3(\gamma''_{vB}) + \phi_4(\gamma''_{vB}, \gamma''_{cB}) < t\phi_1(\gamma_{vA}) + (1-t)\phi_1(\gamma'_{vA}) + t\phi_2(\gamma_{vA}, \gamma_{cA}) + \\
(1-t)\phi_2(\gamma'_{vA}, \gamma'_{cA}) + t\phi_3(\gamma_{vB}) + (1-t)\phi_3(\gamma'_{vB}) + t\phi_4(\gamma_{vB}, \gamma_{cB}) + (1-t)\phi_4(\gamma'_{vB}, \gamma'_{cB}) = \\
t\left(\phi_1(\gamma_{vA}) + \phi_2(\gamma_{vA}, \gamma_{cA}) + \phi_3(\gamma_{vB}) + \phi_4(\gamma_{vB}, \gamma_{cB})\right) + \\
(1-t)\left(\phi_1(\gamma'_{vA}) + \phi_2(\gamma'_{vA}, \gamma'_{cA}) + \phi_3(\gamma'_{vB}) + \phi_4(\gamma'_{vB}, \gamma'_{cB})\right) \iff \\
f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma''}) < tf_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma}) + (1-t)f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma'})
\end{gathered}$$

Since the previous inequality holds for any $t \in (0,1)$ and $\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathbb{R}^4_+$, $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}'$, we have proven that $f(.)$ is strictly concave. □

A corollary of Lemma 2 is that, for any $\bar{\boldsymbol{\gamma}}$, $f_{\bar{\boldsymbol{\gamma}}}(.)$ is strictly quasiconcave. Thus, $f_{\bar{\boldsymbol{\gamma}}}(.)$ is a continuous, strictly quasiconcave function, and the feasible set of attention choices $\boldsymbol{\gamma^i}$ is compact and convex. This proves that there exists a unique solution to the attention allocation problem for any player $i \in [0,1]$ when they face a market in which all other investors have chosen $\bar{\boldsymbol{\gamma}} >> 0$.

Symmetry and uniqueness of equilibrium

The existence of a symmetric equilibrium was already proven in Proposition 1. The fact that any equilibrium consistent with the hypotheses of this proposition must necessarily be symmetric follows trivially from the uniqueness of the solution to the optimal attention allocation problem. Every investor i faces the same market precision matrix $\bar{\Lambda}$, and this is the only aspect of other players' strategies which affects their attention allocation problem. Since they are symmetric in every other respect, this implies that all investors face the same problem, and choose the same unique optimal attention allocation $\boldsymbol{\gamma}^*$. The equilibrium is thus necessarily symmetric.

We thus have that, for all $i \in [0,1]$, $\boldsymbol{\gamma^i} = \bar{\boldsymbol{\gamma}} = \boldsymbol{\gamma} = (\gamma_{vA}, \gamma_{cA}, \gamma_{vB}, \gamma_{cB})^T$. As a result, equations (32)-(35) simplify to:

$$\frac{(N_A - 1)\, p'(\gamma_{vA})}{q_a \left[p(\gamma_{vA})\right]^2 + p(\gamma_{vA})} + \frac{p'(\gamma_{vA})p(\gamma_{cA})}{p(\gamma_{vA})\left[q_a p(\gamma_{vA})p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})\right]} = N_A \lambda \tag{74}$$

$$\frac{N_A p'(\gamma_{cA})p(\gamma_{vA})}{p(\gamma_{cA}\left[q_a p(\gamma_{vA})p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})\right])} = \lambda \tag{75}$$

$$\frac{(N_B - 1)\, p'(\gamma_{vB})}{q_a \left[p(\gamma_{vB})\right]^2 + p(\gamma_{vB})} + \frac{p'(\gamma_{vB})p(\gamma_{cB})}{p(\gamma_{vB})\left[q_a p(\gamma_{vB})p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})\right]} = N_B \lambda \tag{76}$$

$$\frac{N_B p'(\gamma_{cB})p(\gamma_{vB})}{p(\gamma_{cB}\left[q_a p(\gamma_{vB})p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})\right])} = \lambda \tag{77}$$

Since the LHS of each of these equations represents a partial derivative of $f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i})$ evaluated at $\boldsymbol{\gamma^i} = \bar{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$, we can condense them into:

$$\nabla f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) = \lambda \begin{pmatrix} N_A \\ 1 \\ N_B \\ 1 \end{pmatrix} \tag{78}$$

In what follows we prove that, given parameters $N_A, N_B \in \mathbb{N}$ and $\kappa \in \mathbb{R}_{++}$, there is at most one[22] point $\boldsymbol{\gamma}^* \in \mathbb{R}^4_{++}$ such that equation (78) holds and the attention restriction:

$$N_A\gamma^*_{vA} + \gamma^*_{cA} + N_B\gamma^*_{vB} + \gamma^*_{cB} \leq \kappa$$

is satisfied with equality.[23]

Lemma 3. *There is at most one equilibrium in the attention allocation stage in which all players are choosing an interior solution.*

It is convenient to first prove two intermediate results, given by Sublemmas 3.1 and 3.2. To this end, we firstly define, for every $\boldsymbol{\gamma^i} = (\gamma^i_{vA}, \gamma^i_{cA}, \gamma^i_{vB}, \gamma^i_{cB})^T \in \mathbb{R}^4$, a pair of two-dimensional vectors $\boldsymbol{\gamma^i_A} \equiv (\gamma^i_{vA}, \gamma^i_{cA})^T$ and $\boldsymbol{\gamma^i_B} \equiv (\gamma^i_{vB}, \gamma^i_{cB})^T$; note that $\boldsymbol{\gamma^i} = \begin{pmatrix} \boldsymbol{\gamma^i_A} \\ \boldsymbol{\gamma^i_B} \end{pmatrix}$. Additionally, we make use of the fact that $f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i})$, for any $\bar{\boldsymbol{\gamma}}, \boldsymbol{\gamma^i} >> 0$, satisfies:

$$f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i}) = f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{A}}}(\boldsymbol{\gamma^i_A}) + f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{B}}}(\boldsymbol{\gamma^i_B})$$

where, for $m \in \{A, B\}$, $f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}} : \mathbb{R}^2 \to \mathbb{R}$ is given by:

$$f_{\bar{\boldsymbol{\gamma}}_m}(\boldsymbol{\gamma^i_m}) = (N_m - 1)\, ln\left(1 + \frac{p(\gamma^i_{vm})}{q_a\left[p(\bar{\gamma}_{vm})\right]^2}\right) + ln\left(1 + \left(\frac{p(\gamma^i_{vm})p(\gamma^i_{cm})}{p(\gamma^i_{cm}) + N_m p(\gamma^i_{vm})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm}) + N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2\right)$$

Note that $\nabla f_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{\gamma^i}) = \begin{pmatrix} \nabla f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{A}}}(\boldsymbol{\gamma^i_A}) \\ \nabla f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{B}}}(\boldsymbol{\gamma^i_B}) \end{pmatrix}$.

Sublemma 3.1. *For any $m \in \{A, B\}$ and $\kappa_m \in \mathbb{R}_{++}$, there at most one $\boldsymbol{\gamma^*_m}$ such that $\nabla f_{\boldsymbol{\gamma^*}_m}(\boldsymbol{\gamma^*}_m) = \lambda \begin{pmatrix} N_m \\ 1 \end{pmatrix}$ and $N_m\gamma^*_{vm} + \gamma^*_{cm} = \kappa_m$.*

Proof:

We can rearrange equations (74)-(77) and write them as:

$$\frac{p'(\gamma_{vm})}{p(\gamma_{vm})}\left(\frac{(N_m - 1)}{q_a p(\gamma_{vm}) + 1} + \frac{p(\gamma_{cm})}{q_a p(\gamma_{vm})p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right) = N_m\lambda \tag{79}$$

[22] Recall that we already know that a symmetric equilibrium exists. Proving that there is at most one such point with interior solutions then finishes the proof that there is a unique equilibrium in which each investor plays an interior solution. There is always the possibility that, for some model parameters, there is no interior solution equilibrium, and thus there is no $\boldsymbol{\gamma}^*$ which satisfies equation (78). Our model focuses on the other cases.

[23] Note that the attention restriction must be satisfied with equality for any player's best response, and thus for all players in any equilibrium, because $f_{\bar{\boldsymbol{\gamma}}}(.)$ is strictly increasing in each of its arguments (for any $\bar{\boldsymbol{\gamma}}$).

$$\frac{p'(\gamma_{cm})}{p(\gamma_{cm})}\left(\frac{N_m p(\gamma_{vm})}{q_a p(\gamma_{vm})p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right) = \lambda \tag{80}$$

for $m \in \{A, B\}$. Note that this system of two equations is equivalent to $\nabla f_{\boldsymbol{\gamma_m}}(\boldsymbol{\gamma_m}) = \lambda \begin{pmatrix} N_m \\ 1 \end{pmatrix}$ and implies $\frac{\frac{\partial f_{\boldsymbol{\gamma_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma_m})}{\frac{\partial f_{\boldsymbol{\gamma_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma_m})} = N_m$.

The LHS in equation (79) is strictly decreasing in γ_{vm}, as both $\frac{p'(\gamma_{vm})}{p(\gamma_{vm})}$ and the factor between parentheses are decreasing in γ_{vm}, and is also strictly increasing in γ_{cm}. Conversely, the LHS in equation (80) is strictly decreasing in γ_{cm} and strictly increasing in γ_{vm}. As a result of this, if $\gamma'_{vm} < \gamma''_{vm}$ and $\gamma'_{cm} > \gamma''_{cm}$ for two vectors $\boldsymbol{\gamma'_m}, \boldsymbol{\gamma''_m} >> 0$, then $\frac{\frac{\partial f_{\boldsymbol{\gamma'_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma'_m})}{\frac{\partial f_{\boldsymbol{\gamma'_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma'_m})} < \frac{\frac{\partial f_{\boldsymbol{\gamma''_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma''_m})}{\frac{\partial f_{\boldsymbol{\gamma''_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma''_m})}$; clearly, the reverse inequality holds in the opposite case. Consider now a vector $\boldsymbol{\gamma^*_m}$ such that $\frac{\frac{\partial f_{\boldsymbol{\gamma^*_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma^*_m})}{\frac{\partial f_{\boldsymbol{\gamma^*_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma^*_m})} = N_m$ and $N_m \gamma^*_{vm} + \gamma^*_{cm} = \kappa_m$. Since any other vector $\boldsymbol{\gamma_m} \neq \boldsymbol{\gamma^*_m}$ such that $N_m \gamma_{vm} + \gamma_{cm} = \kappa_m$ satisfies $\gamma_{vm} > \gamma^*_{vm}$ and $\gamma_{cm} < \gamma^*_{cm}$ or vice versa, it follows that $\frac{\frac{\partial f_{\boldsymbol{\gamma_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma_m})}{\frac{\partial f_{\boldsymbol{\gamma_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma_m})} \neq \frac{\frac{\partial f_{\boldsymbol{\gamma^*_m}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma^*_m})}{\frac{\partial f_{\boldsymbol{\gamma^*_m}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma^*_m})} = N_m$. Therefore, $\nabla f_{\boldsymbol{\gamma_m}}(\boldsymbol{\gamma_m}) \neq \lambda \begin{pmatrix} N_m \\ 1 \end{pmatrix}$ for all $\lambda \in \mathbb{R}$ and all $\boldsymbol{\gamma_m} \neq \boldsymbol{\gamma^*_m}$ such that $N_m \gamma_{vm} + \gamma_{cm} = \kappa_m$, which proves the intended intermediate result. □

Sublemma 3.2. *For $m \in \{A, B\}$, an increase in κ_m results in a decrease in both elements of $\nabla f_{\boldsymbol{\gamma^*_m}}(\boldsymbol{\gamma^*_m})$ for the new vector $\boldsymbol{\gamma^*_m}$ satisfying the condition from Sublemma 3.1, as long as such a vector continues to exist for the new level of κ_m.*

Proof:

Assume, by way of contradiction, that for $\kappa'_m, \kappa''_m \in \mathbb{R}_{++}$, $\kappa'_m > \kappa''_m$, we have that both of their corresponding unique attention vectors $\boldsymbol{\gamma'_m}, \boldsymbol{\gamma''_m} \in \mathbb{R}^4$ satisfying the conditions from Lemma **??** exist, and one element of $\nabla f_{\boldsymbol{\gamma'_m}}(\boldsymbol{\gamma'_m})$ is greater than or equal to its corresponding element in $\nabla f_{\boldsymbol{\gamma''_m}}(\boldsymbol{\gamma''_m})$. This implies that $\nabla f_{\boldsymbol{\gamma'_m}}(\boldsymbol{\gamma'_m}) \geq \nabla f_{\boldsymbol{\gamma''_m}}(\boldsymbol{\gamma''_m})$, since the quotient of both elements is equal to N_m for both vectors. We also have $\boldsymbol{\gamma'_m} >> \boldsymbol{\gamma''_m}$, since it cannot be the case that $\gamma'_{vm} > \gamma''_{vm}$ and $\gamma'_{cm} < \gamma''_{cm}$, or vice versa, because of the same reasoning involved in proving Lemma (**??**), and if $\boldsymbol{\gamma'_m} \leq \boldsymbol{\gamma''_m}$ then $N_m \gamma'_{vm} + \gamma''_{cm} \leq \kappa''_m < \kappa'_m$.

As a brief detour, consider now $\nabla f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}}(\boldsymbol{\gamma^i_m})$ for two (possibly different) positive vectors $\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}$ and $\boldsymbol{\gamma^i_m}$. The elements in this gradient are given by the LHS of one of the two pairs of first-order condition in equations (32)-(35), which we now write for a generic $m \in \{A, B\}$:

$$\frac{\partial f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}}}{\partial \gamma_{vm}^{i}}(\boldsymbol{\gamma}_{\boldsymbol{m}}^{\boldsymbol{i}}) = \frac{\frac{(N_m-1)p'(\gamma''_{vm})}{q_a[p(\bar{\gamma}_{vm})]^2}}{\left(1+\frac{p(\gamma''_{vm})}{q_a[p(\bar{\gamma}_{vm})]^2}\right)} + \frac{\left(\frac{p'(\gamma''_{vm})[p(\gamma''_{cm})]^2}{[p(\gamma''_{cm})+N_m p(\gamma''_{vm})]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2}{1+\left(\frac{p(\gamma''_{vm})p(\gamma''_{cm})}{p(\gamma''_{cm})+N_m p(\gamma''_{vm})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2} =$$

$$\frac{(N_m-1)p'(\gamma''_{vm})}{q_a\left[p(\bar{\gamma}_{vm})\right]^2+p(\gamma''_m)} + \frac{p'(\gamma''_{vm})[p(\gamma''_{cm})]^2}{q_a\left[p(\gamma''_{cm})+N_m p(\gamma''_{vm})\right]^2\left(\frac{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}\right)^2+\left[p(\gamma''_{cm})+N_m p(\gamma''_{vm})\right]p(\gamma''_{vm})p(\gamma''_{cm})}$$

and

$$\frac{\partial f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}}}{\partial \gamma_{cm}^{i}}(\boldsymbol{\gamma}_{\boldsymbol{m}}^{\boldsymbol{i}}) = \frac{\left(\frac{N_m p'(\gamma''_{cm})[p(\gamma''_{vm})]^2}{[p(\gamma''_{cm})+N_m p(\gamma''_{vm})]^2}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2}{1+\left(\frac{p(\gamma''_{vm})p(\gamma''_{cm})}{p(\gamma''_{cm})+N_m p(\gamma''_{vm})}\right)\frac{1}{q_a}\left(\frac{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}\right)^2} =$$

$$\frac{N_m p'(\gamma''_{cm})[p(\gamma''_{vm})]^2}{q_a\left[p(\gamma''_{cm})+N_m p(\gamma''_{vm})\right]^2\left(\frac{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}\right)^2+\left[p(\gamma''_{cm})+N_m p(\gamma''_{vm})\right]p(\gamma''_{vm})p(\gamma''_{cm})}$$

Notice that $[p(\bar{\gamma}_{vm})]^2$ is strictly increasing in $\bar{\gamma}_{vm}$ and $\left(\frac{p(\bar{\gamma}_{vm})p(\bar{\gamma}_{cm})}{p(\bar{\gamma}_{cm})+N_m p(\bar{\gamma}_{vm})}\right)^2$ is strcitly increasing in both $\bar{\gamma}_{vm}$ and $\bar{\gamma}_{cm}$. This means that $\frac{\partial f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}}}{\partial \gamma_{vm}^{i}}(\boldsymbol{\gamma}_{\boldsymbol{m}}^{\boldsymbol{i}})$ and $\frac{\partial f_{\bar{\boldsymbol{\gamma}}_{\boldsymbol{m}}}}{\partial \gamma_{cm}^{i}}(\boldsymbol{\gamma}_{\boldsymbol{m}}^{\boldsymbol{i}})$ are both strictly decreasing in $\bar{\gamma}_{vm}$ and $\bar{\gamma}_{cm}$.

Applying this notion to $\boldsymbol{\gamma}'_{\boldsymbol{m}}$ and $\boldsymbol{\gamma}''_{\boldsymbol{m}}$, we have that $\boldsymbol{\gamma}'_{\boldsymbol{m}} >> \boldsymbol{\gamma}''_{\boldsymbol{m}}$, which we have already shown to hold, implies $\nabla f_{\boldsymbol{\gamma}'_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}}) << \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$. Since we know that $\nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}''_{\boldsymbol{m}}) \leq \nabla f_{\boldsymbol{\gamma}'_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$, we have $\nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}''_{\boldsymbol{m}}) << \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$.

We now define $\mathbf{u} \equiv \boldsymbol{\gamma}'_{\boldsymbol{m}} - \boldsymbol{\gamma}''_{\boldsymbol{m}}$, and define $d : [0,1] \to \mathbb{R}$:

$$d(t) = \mathbf{u}^T \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}\left(\boldsymbol{\gamma}''_{\boldsymbol{m}} + t\mathbf{u}\right)$$

Since $d(0) = \mathbf{u}^T \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}''_{\boldsymbol{m}}), d(1) = \mathbf{u}^T \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$, and the three vectors $\mathbf{u}$, $\nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}''_{\boldsymbol{m}})$ and $\nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$ are composed of strictly positive elements, we have that $\nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}''_{\boldsymbol{m}}) << \nabla f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(\boldsymbol{\gamma}'_{\boldsymbol{m}})$, which we have just shown to be true, implies $d(1) > d(0)$.

However, $d(t)$ is the directional derivative of $f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(.)$ in the direction marked by $\mathbf{u}$, evaluated at $\boldsymbol{\gamma}''_{\boldsymbol{m}} + t\mathbf{u}$. As such, its derivative is given by $\mathbf{u}^T H_{\boldsymbol{\gamma}''_{\boldsymbol{m}}} \mathbf{u}$, where $H_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}$ is the Hessian matrix of $f_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}(.)$. Since it is easy to check that $f_{\boldsymbol{\gamma}_{\boldsymbol{m}}}(.)$, for any $\boldsymbol{\gamma}_{\boldsymbol{m}} >> 0$, necessarily inherits the property of strict concavity from $f_{\boldsymbol{\gamma}}(.)$, we know that $H_{\boldsymbol{\gamma}''_{\boldsymbol{m}}}$ is negative definite. Thus, since we know that $\boldsymbol{\gamma}'_{\boldsymbol{m}} \neq \boldsymbol{\gamma}''_{\boldsymbol{m}}$ and thus $\mathbf{u} \neq \mathbf{0}$, we have that $d'(t) = \mathbf{u}^T H_{\boldsymbol{\gamma}''_{\boldsymbol{m}}} \mathbf{u} < 0$ for all $t \in [0,1]$, which implies $d(1) < d(0)$. We have thus reached a contradiction, and concluded the proof that both elements of $\nabla f_{\boldsymbol{\gamma}^*_{\boldsymbol{m}}}(\boldsymbol{\gamma}^*_{\boldsymbol{m}})$ decrease in response to an increase in κ_m, if $f_{\boldsymbol{\gamma}^*_{\boldsymbol{m}}}(\boldsymbol{\gamma}^*_{\boldsymbol{m}})$ continues to exist. □

Finally, consider an equilibrium vector $\boldsymbol{\gamma}^*$, verifying equation (78) for some $\lambda \in \mathbb{R}$ and $N_A\gamma^*_{vA} + \gamma^*_{cA} + N_B\gamma^*_{vB} + \gamma^*_{cB} = \kappa$. We define $\kappa_A = N_A\gamma^*_{vA} + \gamma^*_{cA}$ and $\kappa_B = N_B\gamma^*_{vB} + \gamma^*_{cB}$. We know that no alternative equilibrium can be reached by reallocating attentions within group $m \in \{A, B\}$, because that would imply leaving κ_m fixed and varying γ_{vm} and γ_{cm}, and we now fromm

Sublemma 3.1 that it is impossible to do that and preserve the property that $\frac{\frac{\partial f_{\boldsymbol{\gamma}}}{\partial \gamma^i_{vm}}(\boldsymbol{\gamma})}{\frac{\partial f_{\boldsymbol{\gamma}}}{\partial \gamma^i_{cm}}(\boldsymbol{\gamma})} = N_m$.

At the same time, we can deduce from Lemma 3.2 that no other equilibrium can be obtained varying the total amount of attention κ_m paid to each group. Doing so would involve increasing the total amount of attention paid to one group κ_m paid to group $m \in \{A, B\}$ and decreasing its counterpart $\kappa_{m'}$, with $m \neq m' \in \{A, B\}$. Without loss of generality, we can assume that this reallocation between groups involves increasing κ_A and decreasing κ_B. By Lemma 3.2 this means that both elements of $\nabla f_{\boldsymbol{\gamma_A}}(\boldsymbol{\gamma_A})$ are strictly lower in the new equilibrium, whereas both elements in $\nabla f_{\boldsymbol{\gamma_B}}(\boldsymbol{\gamma_B})$ are strictly greater, if the new equilibrium exists. Since we know that $\boldsymbol{\gamma}^*$ satisfies equation (78) and hence $\frac{\partial f_{\boldsymbol{\gamma}^*}}{\partial \gamma^i_{cA}}(\boldsymbol{\gamma}^*) = \frac{\partial f_{\boldsymbol{\gamma}^*}}{\partial \gamma^i_{cB}}(\boldsymbol{\gamma}^*)$, this means that $\frac{\partial f_{\hat{\boldsymbol{\gamma}}}}{\partial \gamma^i_{cA}}(\hat{\boldsymbol{\gamma}}) \neq \frac{\partial f_{\hat{\boldsymbol{\gamma}}}}{\partial \gamma^i_{cB}}(\hat{\boldsymbol{\gamma}})$ for any other equilibrium $\hat{\boldsymbol{\gamma}}$, which means that $\hat{\boldsymbol{\gamma}}$ does not in fact correspond to an equilibrium. [24]

We have therefore proven that, given an equilibrium attention allocation $\boldsymbol{\gamma}^*$, no vector $\hat{\boldsymbol{\gamma}}$ can describe an equilibrium if $N_m\hat{\gamma}_{vm} + \hat{\gamma}_{cm} \neq N_m\gamma^*_{vm} + \gamma^*_{cm}$, or if $(\hat{\gamma}_{vm}, \hat{\gamma}_{cm})^T \neq (\gamma^*_{vm}, \gamma^*_{cm})^T$ with $N_m\hat{\gamma}_{vm} + \hat{\gamma}_{cm} = N_m\gamma^*_{vm} + \gamma^*_{cm}$, for any $m \in \{A, B\}$. Thus, in order for $\hat{\boldsymbol{\gamma}}$ to correspond to an equilibrium attention allocation, it must be the case that $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^*$. □

E.4 Proof of Proposition 4

E.4.1 Proof of result 1: $\gamma_{cm} > \gamma_{vm}$ for $m \in \{A, B\}$

From equations (37) and (39) we have that, for $m \in \{A, B\}$:

$$\frac{N_m p'(\gamma_{cm}) p(\gamma_{vm})}{p(\gamma_{cm})\left[q_a p(\gamma_{vm}) p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})\right])} =$$
$$N_m \left(\frac{p(\gamma_{vm})}{p(\gamma_{cm})}\right)^2 \left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right) \left[\frac{p'(\gamma_{vm}) p(\gamma_{cm})}{p(\gamma_{vm})\left[q_a p(\gamma_{vm}) p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})\right])}\right] = \lambda \qquad (81)$$

[24] Although at this point it is unnecessary, clearly it can be similarly shown that $\frac{\partial f_{\hat{\boldsymbol{\gamma}}}}{\partial \gamma^i_{vA}}(\hat{\boldsymbol{\gamma}}) \neq \left(\frac{N_B}{N_A}\right) \frac{\partial f_{\hat{\boldsymbol{\gamma}}}}{\partial \gamma^i_{vB}}(\hat{\boldsymbol{\gamma}})$,which is also sufficient to show that $\tilde{\boldsymbol{\gamma}}$ does not constitute an equilibrium attention allocation.

Equation (81), together with equations (36) and (38), yields:

$$\frac{(N_m-1)\,p'(\gamma_{vm})}{q_a\left[p(\gamma_{vm})\right]^2+p(\gamma_{vm})}+\frac{p'(\gamma_{vm})p(\gamma_{cm})}{p(\gamma_{vm})\left[q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})\right]}=$$
$$N_m^2\left(\frac{p(\gamma_{vm})}{p(\gamma_{cm})}\right)^2\left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right)\left[\frac{p'(\gamma_{vm})p(\gamma_{cm})}{p(\gamma_{vm})\left[q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})\right])}\right]\iff$$
$$\frac{(N_m-1)\,p'(\gamma_{vm})}{p(\gamma_{vm})\left(q_a p(\gamma_{vm})+1\right)}=$$
$$\left(N_m^2\left(\frac{p(\gamma_{vm})}{p(\gamma_{cm})}\right)^2\left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right)-1\right)\left[\frac{p'(\gamma_{vm})p(\gamma_{cm})}{p(\gamma_{vm})\left[q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})\right])}\right]\iff$$
$$\frac{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})}{p(\gamma_{cm})(q_a p(\gamma_{vm})+1)}=\frac{N_m^2\left(\frac{p(\gamma_{vm})}{p(\gamma_{cm})}\right)^2\left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right)-1}{N_m-1}\iff$$
$$\frac{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})}{(q_a p(\gamma_{vm})+1)}=\frac{N_m^2\left(\frac{[p(\gamma_{vm})]^2}{p(\gamma_{cm})}\right)\left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right)-p(\gamma_{cm})}{N_m-1}\iff$$
$$\frac{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})}{p(\gamma_{vm})(q_a p(\gamma_{vm})+1)}=\frac{N_m^2\left(\frac{p(\gamma_{vm})}{p(\gamma_{cm})}\right)\left(\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}\right)-\frac{p(\gamma_{cm})}{p(\gamma_{vm})}}{N_m-1}\iff$$
$$\frac{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})}{p(\gamma_{vm})(q_a p(\gamma_{vm})+1)}=\frac{N_m^2\left(\frac{p(\gamma_{vm})}{p'(\gamma_{vm})}\right)\left(\frac{p'(\gamma_{cm})}{p(\gamma_{cm})}\right)-\frac{p(\gamma_{cm})}{p(\gamma_{vm})}}{N_m-1}\tag{82}$$

Note that the LHS is increasing in $p(\gamma_{cm})$, and therefore in γ_{cm}. As for the RHS, a sufficient condition for it to be decreasing in γ_{cm} is for $\frac{p'(\gamma_{cm})}{p'(\gamma_{cm})}$ to be weakly decreasing in γ_{cm} for all feasible values of γ_{cm}, which is equivalent to the proposition's hypothesis that $p(.)$ is logarithmically concave for $\gamma\in[0,\kappa]$.

Let γ_{vm}^* be the optimal choice of γ_{vm}. If the LHS is lower than the RHS for $\gamma_{cm}=\gamma_{vm}^*$, then γ_{cm} must be higher than γ_{vm}^* for the equality to hold. We can see that, if $\gamma_{cm}=\gamma_{vm}=\gamma_{vm}^*$, then:

$$\frac{q_a(p(\gamma_{vm}^*))^2+(N_m+1)p(\gamma_{vm}^*)}{q_a(p(\gamma_{vm}^*))^2+p(\gamma_{vm}^*)}=1+\frac{N_m p(\gamma_{vm}^*)}{q_a(p(\gamma_{vm}^*))^2+p(\gamma_{vm}^*)}=1+<N_m+1=$$
$$\frac{(N_m+1)(N_m-1)}{N_m-1}=\frac{N_m^2-1}{N_m-1}=\frac{N_m^2\left(\frac{p(\gamma_{vm}^*)}{p(\gamma_{vm}^*)}\right)\left(\frac{p'(\gamma_{vm}^*)}{p'(\gamma_{vm}^*)}\right)-\frac{p(\gamma_{vm}^*)}{p(\gamma_{vm}^*)}}{N_m-1}$$

We have, then, that the LHS of equation (82) is indeed lower than the RHS when $\gamma_{cm}=\gamma_{vm}^*$ for all γ_{vm}^* such that $p(\gamma_{vm}^*)>0$, which is necessarily the case given that we are considering interior solutions. This proves that the optimal choice of γ_{cm} must necessarily be higher than γ_{vm}^*, therefore there is category learning in this model. □

E.4.2 Proof of result 2: If $N_A > N_B$, then $\gamma_{cA} > \gamma_{cB}$

Assume that $N_A > N_B$ and $\gamma_{cA} \leq \gamma_{cB}$. Using equations (37) and (39), we can obtain:

$$\frac{N_A p'(\gamma_{cA}) p(\gamma_{vA})}{p(\gamma_{cA}) \left[q_a p(\gamma_{vA}) p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})\right])} = \frac{N_B p'(\gamma_{cB}) p(\gamma_{vB})}{p(\gamma_{cB}) \left[q_a p(\gamma_{vB}) p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})\right])} \iff$$

$$\frac{p'(\gamma_{cB})/p(\gamma_{cB})}{p'(\gamma_{cA})/p(\gamma_{cA})} = \frac{\frac{N_A p(\gamma_{vA})}{q_a p(\gamma_{vA}) p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})}}{\frac{N_B p(\gamma_{vB})}{q_a p(\gamma_{vB}) p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})}} \iff$$

$$\frac{p'(\gamma_{cB})/p(\gamma_{cB})}{p'(\gamma_{cA})/p(\gamma_{cA})} = \frac{h(\gamma_{vA}, \gamma_{cA}, N_A)}{h(\gamma_{vB}, \gamma_{cB}, N_B)} \tag{83}$$

where we define $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ as follows:

$$h(\gamma_{vm}, \gamma_{cm}, N_m) = \frac{N_m p(\gamma_{vm})}{q_a p(\gamma_{vm}) p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})}$$

Since $p(\gamma)$ is logarithmically concave for $\gamma \in [0, \kappa]$, we have that $\frac{p'(\gamma)}{p(\gamma)}$ is weakly decreasing in γ in the relevant domain of $p(.)$. We assumed that $\gamma_{cA} \leq \gamma_{cB}$, therefore the LHS of equation (83) is lower than or equal to than 1. At the same time, $h(.)$ is strictly decreasing in γ_{cm} and strictly increasing in γ_{vm} and N_m whenever $p(\gamma_{vm}) > 0$ and $p(\gamma_{cm}) > 0$. This means that, for interior solutions, $N_A > N_B$, $\gamma_{vA} \geq \gamma_{vB}$ and $\gamma_{cA} \leq \gamma_{cB}$, taken together, imply that the RHS is strictly greater than 1. Since this is absurd, we have that these three conditions cannot hold simultaneously, thus $N_A > N_B$ and $\gamma_{cA} \leq \gamma_{cB}$ imply $\gamma_{vA} < \gamma_{vB}$.

Note that the conditions that the RHS is lower than or equal to 1 and $N_A > N_B$ jointly imply:

$$\frac{N_A p(\gamma_{vA})}{q_a p(\gamma_{vA}) p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})} \leq \frac{N_B p(\gamma_{vB})}{q_a p(\gamma_{vB}) p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})} \Rightarrow$$

$$\frac{p(\gamma_{vA})}{q_a p(\gamma_{vA}) p(\gamma_{cA}) + p(\gamma_{cA}) + N_A p(\gamma_{vA})} < \frac{p(\gamma_{vB})}{q_a p(\gamma_{vB}) p(\gamma_{cB}) + p(\gamma_{cB}) + N_B p(\gamma_{vB})}$$

We can re-write equations (36) and (38) as:

$$\left(\frac{p'(\gamma_{vm})}{N_m p(\gamma_{vm})}\right) \left[\frac{(N_m - 1)}{q_a p(\gamma_{vm}) + 1} + \frac{p(\gamma_{cm})}{q_a p(\gamma_{vm}) p(\gamma_{cm}) + p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right] = \lambda \iff$$

$$\left(\frac{p'(\gamma_{vm})}{N_m p(\gamma_{vm})}\right) \left[\frac{(N_m - 1)}{q_a p(\gamma_{vm}) + 1} + \frac{p(\gamma_{cm})}{p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + N_m p(\gamma_{vm})}\right] = \lambda \iff$$

$$\left(\frac{p'(\gamma_{vm})}{N_m p(\gamma_{vm})}\right) \left[\frac{(N_m - 1) p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + (N_m - 1) N_m p(\gamma_{vm}) + p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right)}{\left[q_a p(\gamma_{vm}) + 1\right] \left[p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + N_m p(\gamma_{vm})\right]}\right] = \lambda \iff$$

$$\left(\frac{p'(\gamma_{vm})}{N_m p(\gamma_{vm})}\right) \left[\frac{N_m p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + (N_m - 1) N_m p(\gamma_{vm})}{\left[q_a p(\gamma_{vm}) + 1\right] \left[\left(p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + N_m p(\gamma_{vm})\right)\right]}\right] = \lambda \iff$$

$$\left(\frac{p'(\gamma_{vm})}{p(\gamma_{vm})}\right) \left[\frac{p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + (N_m - 1) p(\gamma_{vm})}{\left[q_a p(\gamma_{vm}) + 1\right] \left[\left(p(\gamma_{cm}) \left(q_a p(\gamma_{vm}) + 1\right) + N_m p(\gamma_{vm})\right)\right]}\right] = \lambda \tag{84}$$

Therefore:

$$\left(\frac{p'(\gamma_{vA})}{p(\gamma_{vA})}\right)\left[\frac{p(\gamma_{cA})\left(q_a p(\gamma_{vA})+1\right)+\left(N_A-1\right)p(\gamma_{vA})}{\left[q_a p(\gamma_{vA})+1\right]\left[\left(p(\gamma_{cA})\left(q_a p(\gamma_{vA})+1\right)+N_A p(\gamma_{vA})\right)\right]}\right]=$$
$$\left(\frac{p'(\gamma_{vB})}{p(\gamma_{vB})}\right)\left[\frac{p(\gamma_{cB})\left(q_a p(\gamma_{vB})+1\right)+\left(N_B-1\right)p(\gamma_{vB})}{\left[q_a p(\gamma_{vB})+1\right]\left[\left(p(\gamma_{cB})\left(q_a p(\gamma_{vB})+1\right)+N_B p(\gamma_{vB})\right)\right]}\right] \iff$$
$$\left(\frac{p'(\gamma_{vA})}{p(\gamma_{vA})}\right)\left(\frac{1}{q_a p(\gamma_{vA})+1}\right)\left[1-\frac{p_{vA}}{\left(p(\gamma_{cA})\left(q_a p(\gamma_{vA})+1\right)+N_A p(\gamma_{vA})\right)}\right]=$$
$$\left(\frac{p'(\gamma_{vB})}{p(\gamma_{vB})}\right)\left(\frac{1}{q_a p(\gamma_{vB})+1}\right)\left[1-\frac{p_{vB}}{\left(p(\gamma_{cB})\left(q_a p(\gamma_{vB})+1\right)+N_B p(\gamma_{vB})\right)}\right] \tag{85}$$

Note that $\gamma_{vA}<\gamma_{vB}$ implies $\left(\frac{p'(\gamma_{vA})}{p(\gamma_{vA})}\right)\leq\left(\frac{p'(\gamma_{vB})}{p(\gamma_{vB})}\right)$, since $\frac{p'(\gamma)}{p(\gamma)}$ is weakly decreasing in γ, and that $\frac{p(\gamma_{vA})}{q_a p(\gamma_{vA})p(\gamma_{cA})+p(\gamma_{cA})+N_A p(\gamma_{vA})}<\frac{p(\gamma_{vB})}{q_a p(\gamma_{vB})p(\gamma_{cB})+p(\gamma_{cB})+N_B p(\gamma_{vB})}$ implies $1-\frac{p_{vA}}{(p(\gamma_{cA})(q_a p(\gamma_{vA})+1)+N_A p(\gamma_{vA})}>1-\frac{p_{vB}}{(p(\gamma_{cB})(q_a p(\gamma_{vB})+1)+N_B p(\gamma_{vB})}$. Since, $\gamma_{vA}<\gamma_{vB}$ also implies $\frac{1}{q_a p(\gamma_{vA})+1}>\frac{1}{q_a p(\gamma_{vB})+1}$, we have that the LHS in equation (85) is strictly greater than the RHS, which is absurd.

This proves that $N_A>N_B$ and $\gamma_{cA}\leq\gamma_{cB}$ cannot hold simultaneously, and thus $N_A>N_B$ implies $\gamma_{cA}>\gamma_{cB}$. □

E.4.3 Proof of result 3: Assuming $p(.)$ is a linear function. If $N_A>N_B$, then $\gamma_{vA}>\gamma_{vB}$

Using equation (82), we can arrive at:

$$\frac{[p(\gamma_{cm})]^2}{p(\gamma_{vm})}+\frac{p(\gamma_{cm})(N_m-1)}{q_a p(\gamma_{vm})+1}-N_m\frac{p(\gamma_{vm})}{p'(\gamma_{vm})}p'(\gamma_{cm})=0$$

The only non-negative root of the previous quadratic equation is:

$$p(\gamma_{cm})=\left(\frac{1}{2}\right)\left[\frac{p(\gamma_{vm})}{q_a p(\gamma_{vm})+1}\right]\left[1-N_m+\sqrt{(N_m-1)^2+4N_m\frac{p'(\gamma_{cm})}{p'(\gamma_{vm})}(q_a p(\gamma_{vm})+1)^2}\right] \tag{86}$$

which, assuming that $p'(\gamma)=K\ \forall\ \gamma$, can be re-written as:

$$p(\gamma_{cm})=\left(\frac{1}{2}\right)\left[\frac{p(\gamma_{vm})}{q_a p(\gamma_{vm})+1}\right]\left[1-N_m+\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}\right] \tag{87}$$

At the same time, assuming $p'(\gamma)=K\ \forall\ \gamma$ also allows us to re-write equation (84) as:

$$\left(\frac{1}{p(\gamma_{vm})}\right)\left[\frac{p(\gamma_{cm})\left(q_a p(\gamma_{vm})+1\right)+\left(N_m-1\right)p(\gamma_{vm})}{\left[q_a p(\gamma_{vm})+1\right]\left[\left(p(\gamma_{cm})\left(q_a p(\gamma_{vm})+1\right)+N_m p(\gamma_{vm})\right)\right]}\right]=\lambda \tag{88}$$

Combining equations (87) and (88), we arrive at:

$$\left(\frac{1}{p(\gamma_{vm})}\right)\left[\frac{\left(\frac{1}{2}\right)p(\gamma_{vm})\left[\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m-1\right]}{[q_a p(\gamma_{vm})+1]\left(\frac{1}{2}\right)p(\gamma_{vm})\left[\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m+1\right]}\right]=\lambda \iff$$
$$\left[\frac{\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m-1}{[q_a p(\gamma_{vm})+1]\,p(\gamma_{vm})\left[1+\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m+1\right]}\right]=\lambda \tag{89}$$

Using the fact that the LHS of equation (89) for $m=A$ equals the LHS of equation (89) for $m=B$, we arrive at:

$$\frac{p(\gamma_{vA})}{p(\gamma_{vB})}=\frac{\frac{\sqrt{(N_A-1)^2+4N_A(q_a p(\gamma_{vA})+1)^2}+N_A-1}{[q_a p(\gamma_{vA})+1]\left[\sqrt{(N_A-1)^2+4N_A(q_a p(\gamma_{vA})+1)^2}+N_A+1\right]}}{\frac{\sqrt{(N_B-1)^2+4N_B(q_a p(\gamma_{vB})+1)^2}+N_B-1}{[q_a p(\gamma_{vB})+1]\left[\sqrt{(N_B-1)^2+4N_B(q_a p(\gamma_{vB})+1)^2}+N_B+1\right]}} \iff$$
$$\frac{p(\gamma_{vA})}{p(\gamma_{vB})}=\frac{g(\gamma_{vA},N_A)}{g(\gamma_{vB},N_B)} \tag{90}$$

where we define $g:\mathbb{R}\times\mathbb{N}\to\mathbb{R}$ as follows:

$$g(\gamma_{vm},N_m)=\frac{\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m-1}{[q_a p(\gamma_{vm})+1]\left[\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m+1\right]}$$

We have that $g(.)$ is increasing in N_m, given that it can be re-written as:

$$g(\gamma_{vm},N_m)=\left(\frac{1}{q_a p(\gamma_{vm})+1}\right)\left(\frac{\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m-1}{\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m+1}\right)$$

To see that it is also decreasing in γ_{vm}, note that it can also be re-written as:

$$g(\gamma_{vm},N_m)=\frac{\sqrt{\left(\frac{N_m-1}{q_a p(\gamma_{vm})+1}\right)^2+4N_m}+\frac{N_m-1}{q_a p(\gamma_{vm})+1}}{\sqrt{(N_m-1)^2+4N_m(q_a p(\gamma_{vm})+1)^2}+N_m+1}$$

This means that, if $N_A>N_B$ and $\gamma_{vA}\leq\gamma_{vB}$, then the LHS of equation (90) is less than or equal to 1, whereas the RHS is strictly greater than 1, which is absurd. This proves that $N_A>N_B$ implies $\gamma_{vA}>\gamma_{vB}$. □

E.4.4 Proof of result 4: Assuming $p(.)$ is a linear function. If $N_A > N_B$, expected absolute trade is higher for assets in A than for assets in B

Recall from equation (51) and the subsequent discussion that:

$$E(|X_{n1}|) = \sqrt{\frac{2\left(\frac{tr(M_{mnn})}{N_m} + \frac{1}{q_a}\right)}{\pi}} \tag{91}$$

Corollary 2 of Lemma 1, from the proof of Proposition 2, can then be used to obtain:

$$tr\left(M_m^{-1}\right) = \frac{1}{\frac{1}{p(\gamma_{vm})} + \frac{N_m}{p(\gamma_{cm})}} + \frac{N_m - 1}{\frac{1}{p(\gamma_{vm})}} = \frac{1}{\frac{p(\gamma_{cm}) + N_m p(\gamma_{vm})}{p(\gamma_{vm})p(\gamma_{cm})}} + (N_m - 1)\, p(\gamma_{vm}) =$$
$$\frac{p(\gamma_{vm})p(\gamma_{cm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})} + (N_m - 1)\, p(\gamma_{vm}) = p(\gamma_{vm})\left[\frac{p(\gamma_{cm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})} + (N_m - 1)\right] =$$
$$p(\gamma_{vm})\left[\frac{p(\gamma_{cm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})} + \frac{(N_m - 1)\, p(\gamma_{cm}) + N_m\,(N_m - 1)\, p(\gamma_{vm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right] =$$
$$p(\gamma_{vm})\left[\frac{N_m p(\gamma_{cm}) + N_m\,(N_m - 1)\, p(\gamma_{vm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right]$$

which in turn yields:

$$M_{mnn} = \frac{tr\left(M_m^{-1}\right)}{N_m} = p(\gamma_{vm})\left[\frac{p(\gamma_{cm}) + (N_m - 1)\, p(\gamma_{vm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})}\right] \tag{92}$$

The RHS of equation (92) is increasing in γ_{cm} and N_m whenever the first two arguments are positive. To see that it is also increasing in γ_{vm}, note that deriving it with respect to γ_{vm} yields:

$$\frac{p(\gamma_{cm}) + (N_m - 1)\, p(\gamma_{vm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})} + p(\gamma_{vm})\left[\frac{(N_m - 1)\,(p(\gamma_{cm}) + N_m p(\gamma_{vm})) - N_m\,(p(\gamma_{cm}) + (N_m - 1)p(\gamma_{vm}))}{[p(\gamma_{cm}) + N_m p(\gamma_{vm})]^2}\right] =$$
$$\frac{p(\gamma_{cm}) + (N_m - 1)\, p(\gamma_{vm})}{p(\gamma_{cm}) + N_m p(\gamma_{vm})} - \frac{p(\gamma_{vm})p(\gamma_{cm})}{[p(\gamma_{cm}) + N_m p(\gamma_{vm})]^2} =$$
$$\frac{[p(\gamma_{cm}) + (N_m - 1)\, p(\gamma_{vm})]\,[p(\gamma_{cm}) + N_m p(\gamma_{vm})] - p(\gamma_{vm})p(\gamma_{cm})}{[p(\gamma_{cm}) + N_m p(\gamma_{vm})]^2} =$$
$$\frac{[p(\gamma_{cm})]^2 + N_m(N_m - 1)\,[p(\gamma_{vm})]^2 + (2N_m - 1)p(\gamma_{vm})p(\gamma_{cm}) - p(\gamma_{vm})p(\gamma_{cm})}{[p(\gamma_{cm}) + N_m p(\gamma_{vm})]^2} =$$
$$\frac{[p(\gamma_{cm})]^2 + N_m(N_m - 1)\,[p(\gamma_{vm})]^2 + (2N_m - 2)p(\gamma_{vm})p(\gamma_{cm})}{[p(\gamma_{cm}) + N_m p(\gamma_{vm})]^2} > 0 \tag{93}$$

We then have that the RHS of equation (92) is increasing in γ_{vm}, γ_{cm} and N_m. Since we have already proven that $N_A > N_B$ implies $\gamma_{vA} > \gamma_{vB}$ and $\gamma_{cA} > \gamma_{cB}$ whenever $p(.)$ is linear, we have that $N_A > N_B$ implies $M_{Ann} > M_{Bnn}$ whenever $p(.)$ is linear. By equation (91), we then have that if $p(.)$ is linear, then $N_A > N_B$ implies that expected absolute trade is greater for assets in A than for assets in B. □

E.5 Relationship between κ and attention:

If $p(.)$ is a linear function, then a higher value of κ (all else being equal) is associated with higher values of γ_{cm} and γ_{vm}, for category $m \in \{A, B\}$, in equilibrium, as long as we consider equilibria with interior solutions and attention symmetry.

Proof:

Equations (36) and (38) can be re-written as:

$$\left(\frac{p'(\gamma_{vm})}{p(\gamma_{vm})}\right)\left(\frac{(N_m-1)}{q_a p(\gamma_{vm})+1}+\frac{p(\gamma_{cm})}{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm})}\right)=N_m\lambda \tag{94}$$

The second factor in the product on the LHS negatively depends on $p(\gamma_{vm})$ and hence also on γ_{vm}. Since $\frac{p'(\gamma_{vm})}{p(\gamma_{vm})} = \frac{K}{p(\gamma_{vm})}$ for some $K > 0$, the first factor also depends negatively on γ_{vm}. This means that, for fixed values of γ_{cm} and N_m, γ_{vm} is higher the lower λ is. We also have that, since the LHS is increasing in $p(\gamma_{cm})$ and thus in γ_{cm}, higher values for γ_{cm} imply higher values for λ, if γ_{vm} is left fixed.

Note as well that equations (37) and (39) can be re-written as:

$$\left(\frac{p'(\gamma_{cm})}{p(\gamma_{cm})}\right)\left(\frac{N_m p(\gamma_{vm})}{q_a p(\gamma_{vm})p(\gamma_{cm})+p(\gamma_{cm})+N_m p(\gamma_{vm}))}\right)=\lambda \tag{95}$$

Here, an analogous argument to the one given previously yield that, for fixed values of γ_{vm} and N_m, γ_{cm} is higher the lower λ is. Similar arguments can also be used to establish that, if γ_{cm} and N_m are left fixed, higher values for λ correspond to higher values of γ_{vm}.

We now consider comparative statics, in which κ increases while N_A and N_B remain constant. If κ rises, then the loosening on the restriction in $N_A\gamma_{vA} + \gamma_{cA} + N_B\gamma_{vB} + \gamma_{cB} \leq \kappa$ implies that, if the restriction is met with equality (which is guaranteed by the fact that each investor's objective function $f_{\boldsymbol{\gamma}}(.)$ is strictly increasing in all four arguments), then at least one element in $\gamma_{vA}, \gamma_{cA}, \gamma_{vB}, \gamma_{cB}$ must increase. Let us assume that γ_{cm} increases for some $m \in \{A, B\}$. If γ_{vm} decreases, then λ must be higher, according to equation (94), but also lower according to equation (95), which is absurd. Therefore, γ_{vm} must also increase. An analogous argument yields that, if γ_{vm} increases, the so must γ_{cm}.

We conclude that κ is greater if and only if both γ_{vm} and γ_{cm} are greater for some $m \in \{A, B\}$. Without loss of generality, let us say this holds for $m = A$. Since $p(.)$ is linear, we have that equation (90) holds. If γ_{vB} is lower when κ increases, then the LHS increases, whereas the RHS decreases because $g(\gamma_{vm}, N_m)$ is decreasing in γ_{vm}. Since this is absurd, we have that γ_{vB} must also increase when κ increases, which entails that γ_{cB} increases as well, given the argument made previously. Thus, we can conclude that a looser restriction on aggregate attention increases attention paid to all signals in equilibrium, rather than inducing investors to concentrate more heavily on one signal in particular, or signals for one particular category,

while reducing attention paid to other signals.

Corollary: A higher value of κ (all else equal, and within equilibria with interior solutions and attention symmetry) is associated with a higher value of $\frac{p(\gamma_{cm})}{p(\gamma_{vm})}$.

Proof:

The previous result establishes that a higher value of κ is associated with a higher value of all attentions γ_{vA}, γ_{cA}, for $m \in \{A, B\}$. Given that $p(.)$ is strictly increasing, this implies that a higher value of κ is associated with higher values of $p(\gamma_{vm}), p(\gamma_{cm})$, for $m \in \{A, B\}$. It is thus sufficient to prove that $\frac{p(\gamma_{cm})}{p(\gamma_{vm})}$ is greater if and only if $p(\gamma_{vm})$ is greater, in equilibrium.

Note that the linearity of $p(.)$ implies that equation (87) holds. Note as well that equation (87) can be re-written as:

$$\frac{p(\gamma_{cm})}{p(\gamma_{vm})} = \left(\frac{1}{2}\right)\left[\frac{1}{q_a p(\gamma_{vm}) + 1}\right]\left[1 - N_m + \sqrt{(N_m - 1)^2 + 4N_m(q_a p(\gamma_{vm}) + 1)^2}\right] \iff$$

$$\frac{p(\gamma_{cm})}{p(\gamma_{vm})} = \left(\frac{1}{2}\right)\left[\frac{1 - N_m}{q_a p(\gamma_{vm}) + 1} + \sqrt{\left(\frac{N_m - 1}{q_a p(\gamma_{vm}) + 1}\right)^2 + 4N_m}\right] \tag{96}$$

To see that $\frac{p(\gamma_{cm})}{p(\gamma_{vm})}$ is increasing in $p(\gamma_{vm})$, define $r : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ such that:

$$r(p, N) = \left(\frac{1}{2}\right)\left[\frac{1 - N}{q_a p + 1} + \sqrt{\left(\frac{N - 1}{q_a p + 1}\right)^2 + 4N}\right]$$

and note that, by equation (96), $\frac{p(\gamma_{cm})}{p(\gamma_{vm})} = r(p(\gamma_{vm}), N_m)$. We have that:

$$\frac{\partial r}{\partial p} = \left(\frac{1}{2}\right)\left[\frac{q_a (N - 1)}{(q_a p + 1)^2} + \left(\frac{1}{2\sqrt{\left(\frac{N-1}{q_a p+1}\right)^2 + 4N}}\right)\left(\frac{2q_a (1 - N)}{(q_a p + 1)^3}\right)\right] =$$

$$\left(\frac{1}{2}\right)\left[\frac{q_a (N - 1)}{(q_a p + 1)^2} + \left(\frac{q_a (1 - N)}{(q_a p + 1)^2\sqrt{(N - 1)^2 + 4N (q_a p + 1)^2}}\right)\right] =$$

$$\left(\frac{1}{2}\right)\left(\frac{q_a (N - 1)\left[\sqrt{(N - 1)^2 + 4N (q_a p + 1)^2} - 1\right]}{(q_a p + 1)^2\sqrt{(N - 1)^2 + 4N (q_a p + 1)^2}}\right) >$$

$$\left(\frac{1}{2}\right)\left(\frac{q_a (N - 1)\left[\sqrt{(N - 1)^2} - 1\right]}{(q_a p + 1)^2\sqrt{(N - 1)^2 + 4N (q_a p + 1)^2}}\right) = \left(\frac{1}{2}\right)\left(\frac{q_a (N - 1)(N - 2)}{(q_a p + 1)^2\sqrt{(N - 1)^2 + 4N (q_a p + 1)^2}}\right)$$

$$\geq 0 \text{ if } N \geq 2$$

This proves that higher values of $p(\gamma_{vA})$ are associated with higher values of $\frac{p(\gamma_{cA})}{p(\gamma_{vA})}$. This

concludes the proof, since it has also been shown that higher values of $p(\gamma_{vA})$ are associated with higher values of κ.

References

Abis, Simona, and Laura Veldkamp. 2024. The changing economics of knowledge production. *The Review of Financial Studies* 37 (1): 89–118.

Abrardi, Luca, Luca Colombo, and Pierluigi Tedeschi. 2020. The value of ignoring risk: competition between better informed insurers. Working paper, Polytechnic University of Turin.

Admati, Anat R. 1985. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica: Journal of the Econometric Society* 629–657.

Andrei, Daniel, and Michael Hasler. 2015. Investor attention and stock market volatility. *The review of financial studies* 28 (1): 33–72.

Arifovic, Jasmina, Xue-zhong He, and Lijian Wei. 2022. Machine learning and speed in high-frequency trading. *Journal of economic dynamics and control* 139 104438.

Barber, Brad M, and Terrance Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies* 21 (2): 785–818.

Barberis, Nicholas, and Andrei Shleifer. 2003. Style investing. *Journal of financial economics* 68 (2): 161–199.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp. 2018. Big data in finance and the growth of large firms. *Journal of monetary economics* 97 71–87.

Ben-Rephael, Azi, Zhi Da, and Ryan D Israelsen. 2017. It depends on where you search: Institutional investor attention and underreaction to news. *The review of financial studies* 30 (9): 3009–3047.

Benamar, Hedi, Thierry Foucault, and Clara Vega. 2021. Demand for information, uncertainty, and the response of US Treasury securities to news. *The review of financial studies* 34 (7): 3403–3455.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008 (10): P10008. 10.1088/1742-5468/2008/10/p10008.

Bose, Devdeepta, Henning Cordes, Sven Nolte, Judith Christiane Schneider, and Colin Farrell Camerer. 2022. Decision weights for experimental asset prices based on visual salience. *The review of financial studies* 35 (11): 5094–5126.

Brunnermeier, Markus, Ralf Lamba, and Constantinos Segura-Rodriguez. 2020. Inverse Selection. Working paper, Princeton University.

Chakrabarty, Bidisha, and Pamela C Moulton. 2012. Earnings announcements and attention constraints: The role of market design. *Journal of accounting and economics* 53 (3): 612–634.

Chiah, Mardy, and Angel Zhong. 2020. Trading from home: The impact of COVID-19 on trading volume around the world. *Finance research letters* 37 101784.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. In search of attention. *The journal of finance* 66 (5): 1461–1499.

Da, Zhi, and Mitch Warachka. 2011. The disparity between long-term and short-term forecasted earnings growth. *Journal of financial economics* 100 (2): 424–442.

Farboodi, Maryam, Adrien Matray, Laura Veldkamp, and Venky Venkateswaran. 2022. Where has all the data gone? *The review of financial studies* 35 (7): 3101–3138.

Farboodi, Maryam, and Laura Veldkamp. 2020. Long-run growth of financial data technology. *American economic review* 110 (8): 2485–2523.

Fedyk, Anastassia. 2021. Disagreement after news: Gradual information diffusion or differences of opinion? *The review of asset pricing studies* 11 (3): 465–501.

Gargano, Antonio, and Alberto G Rossi. 2018. Does it pay to pay attention? *The review of financial studies* 31 (12): 4595–4649.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of economic literature* 57 (3): 535–74. 10.1257/jel.20181020.

Glode, Vincent, and Guillermo Ordoñez. 2024. Technological progress and rent seeking. *The review of financial studies* hhae031. 10.1093/rfs/hhae031.

Goldstein, Itay, Chester S Spatt, and Mao Ye. 2021. Big data in finance. *The review of financial studies* 34 (7): 3213–3225.

Grossman, Sanford J, and Joseph E Stiglitz. 1980. On the impossibility of informationally efficient markets. *The american economic review* 70 (3): 393–408.

Hao, Jing, and Xiong Xiong. 2021. Retail investor attention and firms' idiosyncratic risk: Evidence from China. *International review of financial analysis* 74 101675.

Hillert, Alexander, Heiko Jacobs, and Sebastian Müller. 2014. Media makes momentum. *The review of financial studies* 27 (12): 3467–3501. 10.1093/rfs/hhu061.

Hirshleifer, David, Sonya S Lim, and Siew Hong Teoh. 2011. Limited investor attention and stock market misreactions to accounting information. *The review of asset pricing studies* 1 (1): 35–73.

Huang, Allen H., Hui Wang, and Yi Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary accounting research* 40 (2): 806–841. https://doi.org/10.1111/1911-3846.12832.

Judd, Kenneth L. 1985. The law of large numbers with a continuum of iid random variables. *Journal of economic theory* 35 (1): 19–25.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. A rational theory of mutual funds' attention allocation. *Econometrica* 84 (2): 571–626.

Koijen, Ralph SJ, and Motohiro Yogo. 2022. New perspectives on insurance. *The review of financial studies* 35 (12): 5275–5286.

Lee, Charles MC, Paul Ma, and Charles CY Wang. 2015. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of financial economics* 116 (2): 410–431.

Leitner, Georg, Jaspal Singh, Anton van der Kraaij, and Balázs Zsámboki. 2024. The rise of artificial intelligence: benefits and risks for financial stability. *www.ecb.europa.eu*, https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202405_02~58c3ce5246.en.html.

Li, Xian, James A Hendler, and John L Teall. 2016. Investor attention on the social web. *Journal of behavioral finance* 17 (1): 45–59.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. Rational inattention: A review. *Journal of economic literature* 61 (1): 226–273.

Mohammad, Saif M. 2016. 9 - Sentiment analysis: detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, edited by Meiselman, Herbert L. 201–237, Woodhead Publishing, . https://doi.org/10.1016/B978-0-08-100508-8.00009-6.

Mohrschladt, H., and J.C. Schneider. 2021. Idiosyncratic volatility, option-based measures of informed trading, and investor attention. *Review of derivatives research* 24 (3): 197–220. 10.1007/s11147-021-09175-7.

Mondria, Jordi. 2010. Portfolio choice, attention allocation, and price comovement. *Journal of economic theory* 145 (5): 1837–1864.

Mondria, Jordi, and Climent Quintana-Domeque. 2013. Financial contagion and attention allocation. *The economic journal* 123 (568): 429–454.

Mondria, Jordi, Thomas Wu, and Yi Zhang. 2010. The determinants of international investment and attention allocation: Using internet search query data. *Journal of international economics* 82 (1): 85–95.

Ortmann, Regina, Matthias Pelster, and Sascha Tobias Wengerek. 2020. COVID-19 and investor behavior. *Finance research letters* 37 101717. https://doi.org/10.1016/j.frl.2020.101717.

Peng, Lin, and Wei Xiong. 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80 (3): 563–602.

Peyton, Kevin, and Saritha Unnikrishnan. 2023. A comparison of chatbot platforms with the state-of-the-art sentence BERT for answering online student FAQs. *Results in engineering* 17 100856. https://doi.org/10.1016/j.rineng.2022.100856.

Shen, Dehua, Andrew Urquhart, and Pengfei Wang. 2019. Does twitter predict Bitcoin? *Economics letters* 174 118–122.

Sims, Christopher A. 2003. Implications of rational inattention. *Journal of monetary economics* 50 (3): 665–690.

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2009. Information immobility and the home bias puzzle. *The journal of finance* 64 (3): 1187–1215.

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. Information acquisition and under-diversification. *The review of economic studies* 77 (2): 779–805.