Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Optimal attention allocation: picking alpha or betting on beta?

Zuyao Gu, Yun Shi, Tingjin Yan & Yong Zhou

To cite this article: Zuyao Gu, Yun Shi, Tingjin Yan & Yong Zhou (2024) Optimal attention allocation: picking alpha or betting on beta?, Quantitative Finance, 24:11, 1679-1702, DOI: 10.1080/14697688.2024.2423702

To link to this article: https://doi.org/10.1080/14697688.2024.2423702

Quantitative Finance, 2024
Vol. 24, No. 11, 1679–1702, https://doi.org/10.1080/14697688.2024.2423702

Optimal attention allocation: picking alpha or betting on beta?

ZUYAO GU, YUN SHI, TINGJIN YAN * and YONG ZHOU

Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, and School of Statistics and Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, People's Republic of China

(*Received 1 April 2024; accepted 24 October 2024*)

We investigate a problem of attention allocation and portfolio selection with information capacity constraint and return predictability in a multi-asset framework. In a two-phase formulation, the optimal attention strategy maximizes the combined expected alpha payoffs and expected beta payoffs of the portfolio. Return predictors taking extreme values incentivize the investor to learn about them and this leads to competition among information sources for attention. Moreover, the investor trades with varying skills including picking alphas and betting on beta, depending on the magnitude of the related predictors. Our multi-period analysis using reinforcement learning demonstrates time-horizon effects on attention and investment strategies.

Keywords: Attention allocation; Portfolio selection; Bayesian learning; Return predictability; Reinforcement learning

JEL Classifications: D83, C61, G11, G41

1. Introduction

The allocation of attention, a critical yet limited cognitive resource, plays a pivotal role in shaping an investor's decision-making process. In financial markets, where information is abundant and varied, understanding how investors allocate their attention to maximize investment performance is of paramount importance. This paper delves into the intricacies of attention allocation in a multi-asset and multi-period investment context, exploring the nuanced trade-offs investors face between competing strategies and competing signals while dealing with limited financial resources and limited attention.

The study of limited attention, rooted in psychology, traces back to Kahneman (1973), who highlights the need for individuals to divert attention from other tasks to perform specific tasks. Gabaix (2019) provides a comprehensive overview of this literature. For instance, Da *et al.* (2011) use Google's Search Volume Index as a measure of individual investors' attention, finding that increased attention can predict short-term return spikes and subsequent long-term reversals. Ben-Rephael *et al.* (2017) construct an abnormal institutional attention measure, observing that institutional attention responds more quickly to major news compared to retail attention. Chen *et al.* (2022) summarize previous measures for attention, emphasizing their significant role in predicting stock market trends and aiding in asset allocation.

Sims (2003) initiates the theoretical exploration of limited attention in financial markets with his rational inattention theory, employing Shannon entropy from information theory to quantify information capacity constraints. This theory has been extensively applied to understand various market phenomena, including equity return comovements (Peng and Xiong 2006, Mondria 2010), post-earnings announcement drift (Jacobs and Weber 2016, Andrei *et al.* 2023), consumption patterns (Reis 2006), investor behaviors in different economic cycles (Kacperczyk *et al.* 2016), and retail trading (Barber *et al.* 2024). Within investment literature, a prominent theme revolves around specialized learning strategies, especially in studies employing capacity constraints. Noteworthy works by Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk *et al.* (2014, 2016) emphasize the tendency for investors to prioritize specific information sources due to attentional limitations. This inclination persists even in scenarios with correlated asset returns, as highlighted by Van Nieuwerburgh and Veldkamp (2010). Fund managers' shifting focus between macro-level shocks and micro-level idiosyncrasies during different economic cycles is evident in the works of Kacperczyk *et al.* (2014, 2016). Additionally, Gondhi (2023) and Glasserman and Mamaysky (2023)

*Corresponding author. Email: tjyan@fem.ecnu.edu.cn
All the authors have contributed equally to this work

observe a propensity for investors to concentrate on either macro or micro information.

While the capacity constraint approach dominates the literature, an alternative method for modeling limited attention exists: the information cost approach. This approach, exemplified by Veldkamp (2006) and others, considers limited attention as a monetary cost associated with information gathering and processing. Abel *et al.* (2013) incorporate both information costs and transaction costs in Merton's problem, explaining infrequent adjustments in investors' information acquisition and investment behavior. Andrei and Hasler (2020) demonstrate in a single-asset market with quadratic cost that the optimal attention strategy follows a U-shaped pattern in return predictors. However, despite its theoretical convenience, this approach may not fully align with psychological findings on attention as a cognitive resource distinct from financial wealth (Kahneman 1973, Chabris and Simons 2011). Therefore, in this paper, we adhere to the capacity constraint approach.

In our theoretical model, an agent can invest in one risk-free asset and n risky assets with unobservable expected returns. At each point in time, the investor optimally chooses her portfolio, and the amount of attention allocated to different signals that help to predict the asset returns. By paying more attention to some signals, the investor can acquire more accurate information on certain signals and therefore increase her expected investment performance, but at the expense of decreasing her attention to alternative signals. In other words, the investor faces a dynamic trade-off problem of asset and attention allocation, both across time periods and across multiple signal sources.

We tackle the challenge of information acquisition models with multiple sources by providing explicit solutions under simple two-phase cases to clarify the mechanism of balancing the trade-offs mentioned above first. Our findings show both specialized and diversified learning strategies can emerge, depending on the investment opportunities, information capacity level, and the investor's efficiency in interacting with the information environment. This finding contrasts with previous studies that typically showed investors adopting specialized strategies even in the presence of correlated asset returns.† We also discover U-shaped and inverted U-shaped relationships in attention allocation, contingent on the values of predictors, leading to varied investment strategies between passive diversification and active asset picking. To better link our model's findings to risk management practices in portfolio optimization, we define the investor's benefit from learning as the improvements in the objective function resulting from learning. We clarify the crucial role of attention allocation in effective portfolio risk management in response to volatile market conditions.

Jointly modeling information and portfolio choice, our study sheds light on the interconnections between attention strategy and investment strategy. We demonstrate that the investor's optimal attention allocation is aimed at maximizing the expected squared posterior Sharpe ratio of the portfolio. Investor trading patterns oscillate between passive diversification (beta strategy) and active asset picking (alpha strategy), influenced by predictor magnitudes. Aligning with the switch in investment policy, attention focuses more on firm-specific predictors during an alpha strategy and on systematic predictors during a beta strategy. Furthermore, we emphasize a significant feedback effect emanating from the attention strategy to investment, highlighting that it is not only the investment strategy that influences attention allocation. Intensive learning about a predictor encourages the investor to trade more aggressively on that asset, leading to a less diversified portfolio than in scenarios without such focused information acquisition. This feedback loop may be attributed to a tendency towards overconfidence in investors, particularly for information they have painstakingly gathered, a phenomenon well-documented in the literature (Daniel *et al.* 1998, Gervais and Odean 2001).

For a comprehensive analysis, we extend our study to consider the applications of the model to the emerging cryptocurrency market, a multi-period problem, and asset pricing. The findings based on cryptocurrency data reveal similar patterns to the results in stock markets and highlight the switching between specialized and diversified learning strategies and the close interactions between attention and investment strategies. The multi-period problem is computationally challenging due to the high dimensionality and the capacity constraints on the strategy. To address these problems, we utilize *Q-learning* (Bertsekas 2019) in reinforcement learning. The interaction patterns between attention allocation and investment strategy, observed in the single-period setting, persist under the multi-period framework. Additionally, we uncover distinct time-horizon effects on the proposed attention and investment strategy. Specifically, the optimal portfolio exhibits more aggressive positioning in earlier periods compared to the terminal period. This finding aligns with the 'age effects' identified in life-cycle portfolio management problems (Viceira 2001, Cocco *et al.* 2005). Moreover, we provide brief discussions on the impacts of information acquisition on asset prices. We obtain a relation of the Capital Asset Pricing Model (CAPM) in a two-phase economy. Consistent with Andrei *et al.* (2023), our result predicts a positive relation between the investor's attention to a certain asset and that asset's beta on the signal-announcement day. Resolving more uncertainty through allocating attention rewards the investor with a higher market risk premium, leading to a steeper security market line (SML). However, our model has additional findings that increasing attention to the specific predictor or systematic predictor of a certain asset can also affect other assets' betas.‡

† Based on the classic capital asset pricing model, we assume an incomplete market where the number of risky assets is smaller than the random information sources. This setting differs from those in Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk *et al.* (2016), where the number of risky assets equals that of information sources. Due to this formulation, our optimal attention and investment strategies exhibit rich patterns and interesting interconnections.

‡ The essential reason for this difference is that Andrei *et al.* (2023) does not consider capacity constraint as well as the attention strategy for the systematic predictor, while our framework incorporates these two.

The rest of the paper is structured as follows: Section 2 outlines the model and provides analytical solutions for simple two-phase cases. Section 3 conducts an empirical analysis as a demonstration of how attention and investment strategies interact. Section 4 presents sensitivity analyzes to show how parameters like learning efficiency affect optimal attention allocation and investment strategies. Section 5 extends our model to include multi-period cases and discusses market equilibrium. Section 6 concludes with remarks. Appendices contain theoretical proofs and the algorithm for our dynamic setting problem.

2. The model

2.1. *The financial market and the learning process*

We consider a financial market with one risk-free asset and n risky assets. The risk-free asset is in infinitely elastic supply with the interest rate r_f. In this discrete-time economy, the price vector of the n risky assets is denoted by $\mathbf{p}_t = \left(p_{1,t}, p_{2,t}, \ldots, p_{n,t}\right)^\top$.† The vector of excess dollar returns $\mathbf{r}_t := \mathbf{p}_t - (1 + r_f)\mathbf{p}_{t-1}$ is assumed to satisfy:

$$\mathbf{r}_t = \boldsymbol{a} + \boldsymbol{b} g_{m,t} + \boldsymbol{g}_{e,t}, \tag{1}$$

where $g_{m,t}$ represents a systematic return predictor and $\boldsymbol{g}_{e,t} = (g_{1,t}, g_{2,t}, \ldots, g_{n,t})^\top$ is an n-dimensional vector of firm-specific return predictors for the risky assets. These predictor variables are assumed to be random and independent of each other. $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^\top$ is the intercept term, and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)^\top$ denotes the exposures of the risky assets returns to the systematic predictor.

In our setup, the number of risky assets is smaller than the number of random information sources, i.e. the number of the investor's learning targets. This implies an incomplete market and distinguishes our paper from the literature such as Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk *et al.* (2016) where the portfolio and attention strategy share the same dimension. We believe our setup is appropriate as in practice, the information sources are much more abundant than the investment opportunities. More importantly, under this formulation, both specialized and diversified learning strategies can emerge, leading to interesting interconnections between attention and investment strategies. While in Van Nieuwerburgh and Veldkamp (2010), only one of two learning patterns appears as the optimal strategy, depending on the investor's learning technology and utility function.

We define the index set of the risky assets as $\mathcal{I} := \{1, \ldots, n\}$, and define the index set of the return predictors as $\mathcal{J} := \{1, \ldots, n\} \cup \{m\}$. We adopt the setup in Gârleanu and Pedersen (2013), which is based on the empirical evidence on return predictability, and model each return predictor $g_{j,t+1}$ for $j \in \mathcal{J}$ using mean-reverting process:

$$g_{j,t+1} = \phi_j g_{j,t} + \epsilon_{j,t+1}, \tag{2}$$

where ϕ_j determines the persistence of the return predictor $g_{j,t}$. In other words, $1 - \phi_j$ measures the mean-reversion speed of this return predictor. The noise term $\epsilon_{j,t+1}$ follows a normal distribution $N(0, {\omega_j}^2)$. Following Peng (2005), we assume that the investor knows the true values of ϕ_j and ${\omega_j}^2$. The independence of these return predictors seems to rule out the cross-sectional patterns (Gârleanu and Pedersen 2013, Ma and Zhu 2019) but indeed is without loss of generality. As noted by Van Nieuwerburgh and Veldkamp (2010), we can regard the return predictors as independent principal components from the correlated assets. Then the investors actually learn about and invest in these linear combinations of assets.

To better illustrate the timeline of the model, we provide a diagram in figure 1. In general, the time interval $[t, t+1]$ can be divided into two phases: the learning phase and the trading phase. The learning phase is from t to $t+$. It is the period when the investor with limited attention capacity decides how much attention to allocate to each return predictor. At the end of the learning phase, i.e. time $t+$, the investor observes $n+1$ signals about the next-period predictors. Examples of these signals encompass firm announcements, media coverage, and analyst reports. The precision of the signals depends on the attention allocated, which further influences how the investor updates the belief about the future predictors. Based on the updated belief, the investor optimizes and adjusts the portfolio at the trading phase from $t+$ to $t+1$. In the following, we illustrate the above model in detail.

At time $t+$, the investor observes a noisy signal about the next period's component $g_{j,t+1}$ for each $j \in \mathcal{J}$:

$$s_{j,t+} = g_{j,t+1} + z_{j,t+}, \tag{3}$$

where the noise is also normally distributed as $z_{j,t+} \sim N(0, \eta^2_{j,t+})$ and $z_{j,t+}$'s are independent with each other for different j. Equation (3) implies that the signals contain information about the future components but are perturbed by random noises. $\eta^2_{j,t+}$ is related to the investor's attention strategy and this relation will be specified later.

Then the investor updates her belief about $g_{j,t+1}$ using the signals in a Bayesian manner. From equations (2) and (3), we can write the prior distribution of $g_{j,t+1}$ as $g_{j,t+1} \sim N(\phi_j g_{j,t}, \omega_j^2)$ and the conditional distribution of the signal $s_{j,t+}$ as $s_{j,t+} \mid g_{j,t+1} \sim N(g_{j,t+1}, \eta^2_{j,t+})$. We denote the investor's information set (σ-field) at time t and t_+ by $\mathcal{F}_t = \sigma\left(\left(g_{j,t}\right)_{j\in\mathcal{J}}\right)$ and $\mathcal{F}_{t+} = \sigma\left(\left(g_{j,t}\right)_{j\in\mathcal{J}}, \left(s_{j,t+}\right)_{j\in\mathcal{J}}\right)$, respectively. Then $\mathcal{F}_t \subset \mathcal{F}_{t+}$. By the Bayes' rule, the posterior distribution is $g_{j,t+1} \mid s_{j,t+} \sim N(\mathbb{E}_{t+}\left[g_{j,t+1}\right], \mathrm{Var}_{t+}\left[g_{j,t+1}\right])$ with the posterior mean as a weighted average of the signal and the prior, and the posterior variance as a harmonic mean of the signal and the prior:

$$\mathbb{E}_{t+}\left[g_{j,t+1}\right] = \left(\frac{s_{j,t+}}{\eta^2_{j,t+}} + \frac{\phi_j g_{j,t}}{\omega_j^2}\right) \cdot \left(\frac{1}{\eta^2_{j,t+}} + \frac{1}{\omega_j^2}\right)^{-1}, \tag{4}$$

$$\mathrm{Var}_{t+}\left[g_{j,t+1}\right] = \left(\frac{1}{\eta^2_{j,t+}} + \frac{1}{\omega_j^2}\right)^{-1}, \tag{5}$$

† Throughout the paper, we use boldfaces to distinguish matrices (including vectors) from scalars, and all vectors are column vectors. We use subscripts to indicate the individual assets and time periods.

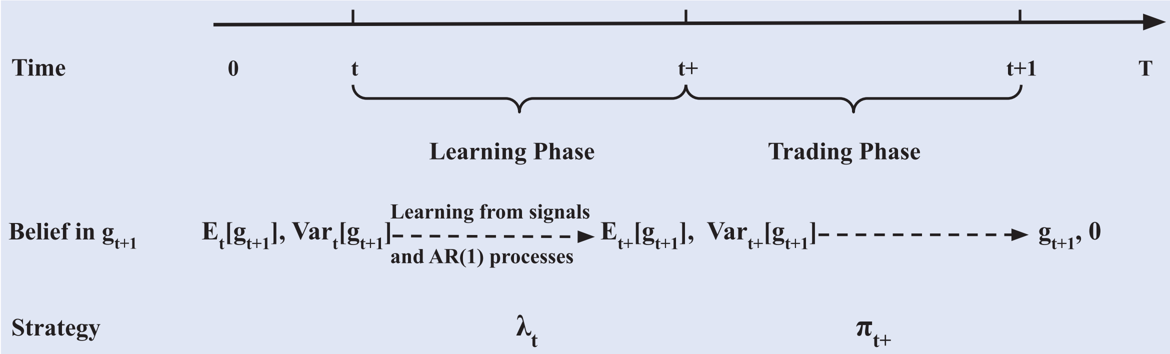


Figure 1. Timeline of the model.

where $\mathbb{E}_{t+}[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t+}]$ and $\mathrm{Var}_{t+}[\cdot] := \mathrm{Var}[\cdot|\mathcal{F}_{t+}]$. To measure the information amount of the signals, we utilize the Shannon differential entropy following Sims (2003) and Peng and Xiong (2006). According to information theory (Shannon 1948, Cover and Thomas 1991), for a continuous random variable X with probability density function $f(\cdot)$, its differential entropy is

$$H(X) := -\mathbb{E}[\log f(X)] = -\int (\log f(x)) f(x)\,\mathrm{d}x.$$

Specifically, for a normally distributed variable X with $X \sim N(\mu, \sigma^2)$, its entropy is given by $H(X) = \frac{1}{2}\log(\sigma^2) + \frac{1}{2}\log(2\pi e)$, which only depends on its variance. For a pair of continuous random variables (X, Y) with joint density function $f(x, y)$ and conditional density function $f(x \mid y)$ of X conditional on Y, the conditional entropy of X given Y is given by

$$H(X \mid Y) := -\int\int f(x, y)\log f(x \mid y)\,\mathrm{d}x\,\mathrm{d}y.$$

Then the mutual information $I(X; Y)$, which is defined as the difference between the marginal entropy of X and the conditional entropy of X given Y, is to quantify the amount of information contained in Y about X.

Given that the prior distribution of $g_{j,t+1}$, $N(\phi_j g_{j,t}, \omega_j^2)$, and the posterior distribution of $g_{j,t+1}$ conditional on $s_{j,t+}$, $N(\mathbb{E}_{t+}[g_{j,t+1}], \mathrm{Var}_{t+}[g_{j,t+1}])$, we apply the difference between the entropy of the prior and the posterior to measure the amount of information contained in $s_{j,t+}$ about $g_{j,t+1}$ by $I(g_{j,t+1}; s_{j,t+})$:

$$\begin{aligned} I(g_{j,t+1}; s_{j,t+}) &= H(g_{j,t+1}) - H(g_{j,t+1} \mid s_{j,t+}) \\ &= \frac{1}{2}\log\left(\frac{\omega_j^2}{\mathrm{Var}_{t+}[g_{j,t+1}]}\right). \end{aligned} \tag{6}$$

The detailed derivations can be found in Appendix A.1. The amount of information contained in the signal is related to the ratio between the prior and posterior variances of the return predictor.

The investor faces a capacity constraint that limits the total amount of information he or she can process. Let $K > 0$ be the information capacity and $\boldsymbol{\lambda}_t = (\lambda_{1,t}, \ldots, \lambda_{n,t}, \lambda_{m,t})^\top$ be the attention strategy satisfying $\sum_{j\in\mathcal{J}} \lambda_{j,t} = 1$. For $j \in \mathcal{J}$, $\lambda_{j,t}$ represents the proportion of K allocated to the next-period predictor $g_{j,t+1}$. With a higher level of $\lambda_{j,t}$, the signal $s_{j,t+}$, which is the output of the investor's information processing, should contain a larger amount of information. Thus, following Peng and Xiong (2006), we assume a positive linear relationship between the amount of attention $\lambda_{j,t}K$ and the quantity of information $I(g_{j,t+1}; s_{j,t+}) = I(g_{j,t+1}; s_{j,t+}|\lambda_{j,t})$:

$$I(g_{j,t+1}; s_{j,t+} \mid \lambda_{j,t}) = \frac{1}{2}\theta_j \lambda_{j,t} K, \tag{7}$$

where θ_j denotes the learning efficiency of $g_{j,t+1}$. A larger θ_j indicates that the investor processes information about the component $g_{j,t+1}$ more efficiently.

Combining equations (5)–(7), we obtain the posterior variance of $g_{j,t+1}$ and the variance of $z_{j,t+}$ as follows

$$\mathrm{Var}_{t+}[g_{j,t+1}] = \omega_j^2 e^{-\theta_j\lambda_{j,t}K}, \tag{8}$$

$$\eta_{j,t+}^2 = \left(\frac{1}{\mathrm{Var}_{t+}[g_{j,t+1}]} - \frac{1}{\omega_j^2}\right)^{-1} = \frac{\omega_j^2}{e^{\theta_j\lambda_{j,t}K} - 1}. \tag{9}$$

The posterior variance of the predictor and the variance of the signal noise decrease with the attention strategy $\lambda_{j,t}$, learning efficiency θ_j as well as the total capacity K. Moreover, as $\lambda_{j,t}, \theta_j$ and K descend to zero, $\eta_{j,t+}^2$ goes to infinity, leading to an uninformative signal. Then equations (4) and (9) imply that the posterior mean of $g_{j,t+1}$ given $s_{j,t+}$ is

$$\mathbb{E}_{t+}[g_{j,t+1}] = \phi_j g_{j,t} + \left(1 - e^{-\theta_j\lambda_{j,t}K}\right)\left(s_{j,t+} - \phi_j g_{j,t}\right). \tag{10}$$

At the trading phase, i.e. from time $t+$ to time $t+1$, the investor constructs a portfolio across assets according to the signals. At time $t+1$, the investor observes the realized value of the components $\boldsymbol{g}_{t+1}$ and obtains the payoffs. Let W_t be the wealth process and $\boldsymbol{\pi}_{t+} = (\pi_{1,t+}, \pi_{2,t+}, \ldots, \pi_{n,t+})^\top$ be the investment strategy with components equal to the shares of risky assets. The investor's wealth process W_t satisfies

$$W_{t+1} = (1 + r_f)W_t + \boldsymbol{\pi}_{t+}^\top \boldsymbol{r}_{t+1}. \tag{11}$$

We define the set of all the admissible attention strategies at time t and the set of all the admissible investment strategies at time $t+$ as follows:†

$$\mathcal{A}_{\lambda,t} = \left\{\boldsymbol{\lambda}_t = (\lambda_{1,t}, \lambda_{2,t}, \ldots, \lambda_{n,t}, \lambda_{m,t})^\top \in [0,1]^{n+1};\right.$$

† It is important to note that the investment strategies $\boldsymbol{\pi}_{t+}$ are not assumed to be static or deterministic functions of the attention

$$\left.\sum_{j\in\mathcal{J}}\lambda_{j,t}=1,\lambda_{j,t}\text{ is }\mathcal{F}_t\text{—measurable for }j\in\mathcal{J}\right\},$$

$$\mathcal{A}_{\pi,t+}=\Big\{\boldsymbol{\pi}_{t+}=(\pi_{1,t+},\pi_{2,t+},\ldots,\pi_{n,t+})^\top \in\mathbb{R}^n,\pi_{i,t+}\text{ is }\mathcal{F}_{t+}\text{—measurable for }i\in\mathcal{I}\Big\}.$$

REMARK 2.1 The proposed attention allocation model can be further extended to incorporate more empirical stylized facts. The interested readers are referred to Andrei and Hasler (2020) and Zhang *et al.* (2022) to see the investigations of attention allocation with stochastic volatility and regime-switching market. Moreover, one may consider other types of return predictability such as cointegration (see, e.g. Ma and Zhu 2019, Yan *et al.* 2022 for applications in the dynamic portfolio problem). Some practical issues can also be formulated in our framework like transaction cost and model uncertainty (see, e.g. Gârleanu and Pedersen 2013, Luo 2017). As our focus is on the interactions between attention and investment strategies in an incomplete market, these interesting investigations are out of the scope of this paper. Moreover, while our model is built upon a representative investor, it is also interesting and practically relevant to consider multiple types of investors. For instance, empirical studies by Barber and Odean (2008) and Liu *et al.* (2023) have shown that institutional investors and retail investors exhibit different attention patterns and trading behaviors due to discrepancies in their attention capacity. Incorporating investor heterogeneity into the theoretical model would provide a deeper understanding of the market equilibrium and asset pricing. We provide a brief discussion on asset pricing when the market includes a representative investor and noise investors in Section 5.3.

2.2. *Investor preference and optimal strategy*

The investor calculates the excess return in the next period, i.e. $W_{t+1}-(1+r_f)W_t$, and optimizes Markowitz's mean-variance criterion of it. The optimization problem includes two steps as described as follows:

$$\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}}\mathbb{E}_t\left[\max_{\boldsymbol{\pi}_{t+}\in\mathcal{A}_{\pi,t+}}\boldsymbol{\pi}_{t+}^\top\mathbb{E}_{t+}\left[\mathbf{r}_{t+1}\right]-\frac{\gamma}{2}\boldsymbol{\pi}_{t+}^\top\operatorname{Var}_{t+}\left[\mathbf{r}_{t+1}\right]\boldsymbol{\pi}_{t+}\right],\tag{12}$$

where $\mathbb{E}_{t+}[\mathbf{r}_{t+1}]$ is the vector of posterior mean of excess returns, $\operatorname{Var}_{t+}[\mathbf{r}_{t+1}]$ is the posterior covariance matrix of excess returns, and $\gamma>0$ is a risk-aversion coefficient.

By combining the model of excess returns (1) and the expressions (10) and (8) for the posterior mean and variance of $g_{j,t+1}$, we can rewrite the posterior mean vector and covariance matrix of excess returns as†

$$\begin{aligned}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)&:=\mathbb{E}_{t+}\left[\mathbf{r}_{t+1}\right]\\&=\mathbb{E}_{t+}\left[\boldsymbol{a}+\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{b}\end{pmatrix}\boldsymbol{g}_{t+1}\right]\\&=\boldsymbol{a}+\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{b}\end{pmatrix}\operatorname{diag}\left[\phi_j g_{j,t}+(1-e^{-\theta_j\lambda_{j,t}K})\right.\\&\qquad\left.(s_{j,t+}-\phi_j g_{j,t})\middle|j\in\mathcal{J}\right]\mathbf{1}_{n+1},\end{aligned}\tag{13}$$

$$\begin{aligned}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)&:=\operatorname{Var}_{t+}\left[\mathbf{r}_{t+1}\right]\\&=\operatorname{Var}_{t+}\left[\boldsymbol{a}+\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{b}\end{pmatrix}\boldsymbol{g}_{t+1}\right]\\&=\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{b}\end{pmatrix}\operatorname{diag}\left[\omega_j^2e^{-\theta_j\lambda_{j,t}K}\middle|j\in\mathcal{J}\right]\begin{pmatrix}\boldsymbol{I}_n\\ \boldsymbol{b}^\top\end{pmatrix},\end{aligned}\tag{14}$$

where $\mathbf{I}_n$ is the $n\times n$ identity matrix and $\mathbf{1}_{n+1}$ is an $n+1$-dimensional vector of ones.

Equation (14) indicates that with more attention allocated to a return predictor, the investor can speculate the corresponding next-period asset return at higher precision. This posterior covariance matrix is independent of the value of signals and only relies on the information at time t. Hence, we adopt the subscript 't' rather than '$t+$'. Note that $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)$ is always positive definite for any $\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}$.

To address the two-layer optimization problem (12), we first solve the inner optimization problem given $\boldsymbol{\lambda}_t$:

$$\max_{\boldsymbol{\pi}_{t+}\in\mathcal{A}_{\pi,t+}}\left[\boldsymbol{\pi}_{t+}^\top\mathbb{E}_{t+}\left[\mathbf{r}_{t+1}\right]-\frac{\gamma}{2}\boldsymbol{\pi}_{t+}^\top\operatorname{Var}_{t+}\left[\mathbf{r}_{t+1}\right]\boldsymbol{\pi}_{t+}\right].$$

From the first-order condition $\mathbb{E}_{t+}[\mathbf{r}_{t+1}]-\gamma\operatorname{Var}_{t+}[\mathbf{r}_{t+1}]\boldsymbol{\pi}_{t+}=\mathbf{0}$, it follows that for any given $\boldsymbol{\lambda}_t$, it is optimal to adopt Markowitz's mean-variance portfolio:

$$\boldsymbol{\pi}_{t+}^{\mathrm{mv}}:=\frac{1}{\gamma}\operatorname{Var}_{t+}\left[\mathbf{r}_{t+1}\right]^{-1}\mathbb{E}_{t+}\left[\mathbf{r}_{t+1}\right]=\frac{1}{\gamma}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t).$$

Moreover, we introduce the posterior Sharpe ratio as the ratio of the posterior mean to the posterior volatility of the return of the portfolio $\boldsymbol{\pi}_{t+}^{\mathrm{mv}}$. Then the posterior Sharpe ratio is a function of $\boldsymbol{\lambda}_t$ given by

$$\begin{aligned}SR_{p,t+}(\boldsymbol{\lambda}_t)&:=\frac{\boldsymbol{\pi}_{t+}^{\mathrm{mv}\top}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)}{\sqrt{\boldsymbol{\pi}_{t+}^{\mathrm{mv}\top}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)\boldsymbol{\pi}_{t+}^{\mathrm{mv}}}}\\&=\sqrt{\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)^\top\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)}.\end{aligned}$$

By substituting the optimal investment strategy $\boldsymbol{\pi}_{t+}^{\mathrm{mv}}=\frac{1}{\gamma}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)$ into the optimization problem (12), the optimal attention problem is reformulated as

$$\begin{aligned}&\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}}\mathbb{E}_t\left[\frac{1}{2\gamma}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)^\top\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)\right]\\&=\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}}\mathbb{E}_t\left[\frac{1}{2\gamma}SR_{p,t+}^2(\boldsymbol{\lambda}_t)\right].\end{aligned}\tag{15}$$

THEOREM 2.1 *The investor's optimal attention allocation problem is*

$$\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}}\mathbb{E}_t\left[\frac{1}{2\gamma}SR_{p,t+}^2(\boldsymbol{\lambda}_t)\right].\tag{16}$$

There exists an optimal attention strategy $\boldsymbol{\lambda}_t^*$ that solves the problem (16). Moreover, the investor's optimal investment strategy is

$$\boldsymbol{\pi}_{t+}^*=\frac{1}{\gamma}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t^*)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t^*).\tag{17}$$

strategy $\boldsymbol{\lambda}_t$. The only requirement is that $\boldsymbol{\pi}_{t+}$ should utilize the information at time $t+$, $\mathcal{F}_{t+}$, which depends on the attention strategies $\boldsymbol{\lambda}_t$ through the selected signals $\mathbf{s}_{t+}=(s_{1,t+},s_{2,t+},\ldots,s_{n,t+},s_{m,t+})^\top$. Once the optimal attention allocation strategies $\boldsymbol{\lambda}_t^*$ are determined, the investor will update the posterior belief and adjust the investment strategies $\boldsymbol{\pi}_{t+}$ accordingly, which will be detailed in the following subsection.

† We use $\operatorname{diag}\left[x_j|j\in\mathcal{J}\right]$ to denote a diagonal matrix with each diagonal element equal to x_j.

Appendix A.2 provides the proof of Theorem 2.1. Theorem 2.1 shows that the attention allocation problem aims at maximizing the portfolio's expected squared posterior Sharpe ratio and the attention problem can be affected by the upcoming investment pattern. To understand this, we rewrite the objective in the attention allocation problem (16) as:

$$\frac{1}{2}\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}} \mathbb{E}_t\left[\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}_t)^\top \boldsymbol{\pi}^{\mathrm{mv}}_{t+}\right]$$
$$= \frac{1}{2}\max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}} \left\{ \underbrace{\mathbb{E}_t\left[\sum_{i=1}^{n}\left(a_i + \mathbb{E}_{t+}\left[g_{i,t+1}\right]\right)\pi^{\mathrm{mv}}_{i,t+}\right]}_{\text{Expected alpha payoffs}} \right.$$
$$\left. + \underbrace{\mathbb{E}_t\left[\sum_{i=1}^{n} b_i \mathbb{E}_{t+}\left[g_{m,t+1}\right]\pi^{\mathrm{mv}}_{i,t+}\right]}_{\text{Expected beta payoffs}} \right\} \tag{18}$$

The above expression indicates that the investor's optimal attention allocation seeks to maximize the combined value of the expected dollar returns at both the firm-specific and the systematic levels with equal weights. The first term represents the alpha return of the strategy, while the second term stands for the beta return of the strategy. When the investor holds an extreme perspective regarding the next-period predictor ($g_{i,t+1}$ or $g_{m,t+1}$), it can influence the investor to attach greater importance to that predictor during the optimization process described by (18). Consequently, the dynamic movements of these return predictors can redirect the investor's focus away from pursuing alpha returns (or beta returns) and towards concentrating on the other, leading to time-varying trading patterns. We will illustrate this in detail in the numerical study later.

On the other hand, the investor's investment strategy is also influenced by her attention allocated to the return predictors. Generally speaking, being attentive decreases the uncertainty of return predictors and improves the precision of the future return. Under the impact of an attention strategy $\boldsymbol{\lambda}^*_t$, the vector of the investor's optimal positions is $\boldsymbol{\pi}^*_{t+} = \frac{1}{\gamma}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}^*_t)^{-1}\boldsymbol{\mu}_{t+}(\boldsymbol{\lambda}^*_t)$ and its average is $\mathbb{E}_t\left[\boldsymbol{\pi}^*_{t+}\right] = \frac{1}{\gamma}\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}^*_t)^{-1}\mathbb{E}_t\left[\boldsymbol{r}_{t+1}\right]$. Consider a diagonal matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)$ (in this case $\mathbf{b}=0$). More attention paid to an asset i will increase the ith diagonal element in precision matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_t)^{-1}$ and motivate the investor to long more or short more that asset (the direction depends on the sign of $\mathbb{E}_t\left[\boldsymbol{r}_{t+1}\right]$). The attention strategy, therefore, leads to an underdiversified portfolio, compared to the well-diversified Markowitz portfolio $\frac{1}{\gamma}\boldsymbol{\Sigma}_t(\mathbf{0})^{-1}\mathbb{E}_t\left[\boldsymbol{r}_{t+1}\right]$ when there is no information acquisition.

Our model incorporates the empirical evidence on the mean-reverting nature of common return predictors (e.g. Poterba and Summers 1988, Gârleanu and Pedersen 2013, Campbell 2017) and the realistic adjustment of portfolio weightings, leading to an attention problem aiming at maximizing the portfolio's squared Sharpe ratio. This differs from Peng and Xiong (2006), in which the authors study investor attention by considering independent fundamentals over time and the portfolio cannot be adjusted. As a result, their attention allocation turns out to be a problem of minimizing the posterior variance of the next-period payoff. In our model, the level of return predictors becomes an important quantity in the investor's decision-making.

In addition, the optimization problem (16) is generally neither concave nor convex. We show in Section 2.3 that both corner solutions and interior solutions can emerge in specific scenarios. In other words, the optimal attention strategy may exhibit different patterns including specialized learning and diversified learning. This is different from the result in Van Nieuwerburgh and Veldkamp (2010) which implies only corner solutions under the mean-variance criterion. The reason is that in our incomplete-market formulation, the investor cannot trade purely a single predictor and then diversified learning is not always dominated by specialized learning especially when the information capacity is sufficiently large. While Van Nieuwerburgh and Veldkamp (2010) considers a complete market with a one-to-one correspondence between assets and learning sources and the investor specializes in learning about one highest-information-value asset.

2.3. Analytical results in specific cases

In this subsection, by considering specific cases, we present analytical results of the optimal attention and investment strategy and provide the intuitions behind them. We first assume there is one risky asset with two return predictors in Theorem 2.2. Next, we further investigate a scenario including two risky assets without systematic predictor in Theorem 2.4.

THEOREM 2.2 *Suppose that there is only one risky asset with two return-predictive fundamentals $g_{1,t}$ and $g_{m,t}$, the investor's optimal attention strategy is given by:*

(i) *(Specialized learning) if $\theta_1\omega_1^2 e^{-\theta_1 K} > \theta_m b^2\omega_m^2$, then $\lambda^*_{1,t}=1$ and $\lambda^*_{m,t}=0$;*

(ii) *(Diversified learning) if $e^{-\theta_m K} \leqslant \frac{\theta_1\omega_1^2}{\theta_m b^2\omega_m^2} \leqslant e^{\theta_1 K}$, then*

$$\lambda^*_{1,t} = \frac{\theta_m}{\theta_1+\theta_m} + \frac{1}{(\theta_1+\theta_m)K}\ln\frac{\theta_1\omega_1^2}{b^2\theta_m\omega_m^2},$$
$$\lambda^*_{m,t} = \frac{\theta_1}{\theta_1+\theta_m} - \frac{1}{(\theta_1+\theta_m)K}\ln\frac{\theta_1\omega_1^2}{b^2\theta_m\omega_m^2};$$

(iii) *(Specialized learning) if $\theta_1\omega_1^2 < \theta_m b^2\omega_m^2 e^{-\theta_m K}$, then $\lambda^*_{1,t}=0$ and $\lambda^*_{m,t}=1$.*

Moreover, the optimal investment strategy is

$$\pi^*_{t+} = \frac{1}{\gamma}\left(\omega_1^2 e^{-\theta_1\lambda^*_{1,t}K} + b^2\omega_m^2 e^{-\theta_m\lambda^*_{m,t}K}\right)^{-1}$$
$$\times\left\{a + \phi_1 g_{1,t} + \left(1-e^{-\theta_1\lambda^*_{1,t}K}\right)\left(s_{1,t+}-\phi_1 g_{1,t}\right)\right.$$
$$\left. + b\left[\phi_m g_{m,t} + \left(1-e^{-\theta_m\lambda^*_{m,t}K}\right)\left(s_{m,t+}-\phi_m g_{m,t}\right)\right]\right\}.$$

Although there is only one asset, the investor's learning strategy can be diversified between two predictors. In this case, there is no need to differentiate the firm-specific predictor and the systematic predictor but only regard them as predictors with different loadings: 1 for $g_{1,t}$ and b for $g_{m,t}$. The proofs of Theorem 2.2 in Appendix A.2 indicate that the optimal attention allocation problem of maximizing the expected squared Sharpe ratio in Theorem 2.1 is reduced to

$$\begin{aligned}&\min_{(\lambda_{1,t},\lambda_{m,t})\in\mathcal{A}_{\lambda,t}} \operatorname{Var}_{t+}[r_{t+1}]\\ &= \min_{(\lambda_{1,t},\lambda_{m,t})\in\mathcal{A}_{\lambda,t}} \operatorname{Var}_{t+}\left[g_{1,t+1}\right] + b^2 \operatorname{Var}_{t+}\left[g_{m,t+1}\right]\\ &= \min_{(\lambda_{1,t},\lambda_{m,t})\in\mathcal{A}_{\lambda,t}} \omega_1^2 e^{-\theta_1\lambda_{1,t}K} + b^2\omega_m^2 e^{-\theta_m\lambda_{m,t}K}. \end{aligned} \tag{19}$$

Therefore, the levels of $g_{1,t}, g_{m,t}$ do not affect the attention strategy in this single-asset case, which is a similar result to that of Peng and Xiong (2006). This is plausible since there is only one asset, and the investor always holds both two predictors with a constant ratio of 1:b. In other words, the investor can not change the portfolio weights of two predictors. As a result, the investor ignores the magnitudes of two predictors and chooses to optimize the total risk measured by the return variance (19).

Interestingly, Theorem 2.2 indicates that both specialized learning and diversified learning can emerge and they are closely related to the investor's information capacity K. Keeping all the remaining parameters fixed, when K is large, the interval in the second scenario has a wide range, indicating the investor is more likely to adopt the way of diversified learning. Conversely, when K is small, the investor tends to be a specialist by prioritizing a single predictor due to the capacity limit.

Moreover, in this single-asset case, the investor prefers to learn the return predictor with a larger prior uncertainty (i.e. w_1^2 and w_m^2) and a higher loading (i.e. 1 and b). However, the effect of learning efficiency (θ_1 and θ_2) appears to be subtle. In figure 2, we provide a diagram to illustrate the relationship between the learning pattern and the learning efficiency. The model parameters are chosen to have similar scales to the estimates based on the daily price of Vanguard Utilities Index Fund (VPU) (more details are given in the empirical study of Section 3). Specifically, we choose $b = 0.5$, $\omega_1^2 = 0.001$, $\omega_m^2 = 0.002$ and $K = 1$. As depicted in figure 2, for a broad range of learning efficiencies, diversified learning is likely to occur when both of two learning efficiencies θ_1 and θ_m are relatively large, while specialized learning happens when one of θ_1, θ_m dominates the other. Suppose we fix θ_m to be 0.5 as indicated by the horizontal dashed line. As θ_1 increases, the investor switches between specialized learning and diversified learning frequently. To be specific, when θ_1 is very small (less than 0.15), due to the low efficiency of learning about $g_{1,t+1}$, the investor will focus exclusively on acquiring information about $g_{m,t+1}$. As θ_1 increases, the investor starts to allocate attention to the predictor $g_{1,t+1}$ and becomes a diversified learner. When θ_1 continues to increase (greater than approximately 0.4), the investor finds it optimal to be a specialist again but this time is to learn about $g_{1,t+1}$. Ultimately, when learning about $g_{1,t+1}$ at an extremely high-efficiency level (θ_1 exceeds approximately 2.25), the investor once again adopts a diversified learning strategy. In this case, a part of the attention is sufficient to reduce the uncertainty of $g_{1,t+1}$ to a satisfactory level. Consequently, the investor also pays attention to $g_{m,t+1}$ to further mitigate the overall uncertainty of asset returns.

Theorem 2.2 has additional predictions on the investment strategy. We calculate the expected position over the signals

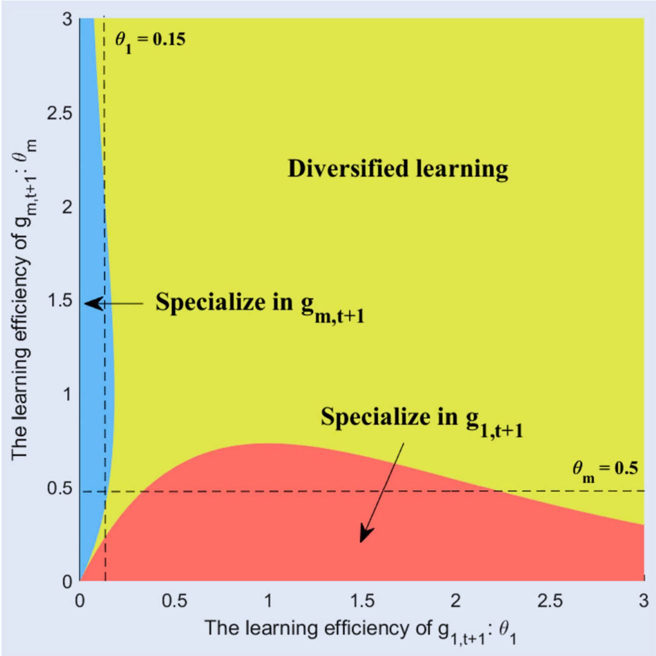


Figure 2. The optimal attention allocation in Theorem 2.2 with varying learning efficiency. We assume the market has only one risky asset with two return predictors. The blue domain represents that the investor only pays attention to the signal about the low-exposure predictor; the yellow domain means that the investor generally learns about both predictors; the red domain represents that the investor specializes in the high-exposure predictor. Here, $b = 0.5$, $\omega_1^2 = 0.001$, $\omega_m^2 = 0.002$ and $K = 1$.

$$\mathbb{E}_t\left[\pi_{t+}^*\right] = \frac{1}{\gamma}\left(\omega_1^2 e^{-\theta_1\lambda_{1,t}^*K} + b^2\omega_m^2 e^{-\theta_m\lambda_{m,t}^*K}\right)^{-1}\mathbb{E}_t\left[r_{t+1}\right].$$

This implies the information acquisition on return predictors will result in a reduced posterior variance in the denominator of the expected position. Consequently, it leads to an underdiversified portfolio, which means more aggressive trading, on average. The direction of the trading depends on the signs of the expected return.

We define the investor's *benefit from learning* as the improvements of the objective function (16) through information acquisition. Specifically, it is given by

$$\Delta(\mathbf{g}_t) := \max_{\boldsymbol{\lambda}_t\in\mathcal{A}_{\lambda,t}} \mathbb{E}_t\left[\frac{1}{2\gamma}SR_{p,t+}^2(\boldsymbol{\lambda}_t)\right] - \mathbb{E}_t\left[\frac{1}{2\gamma}SR_{p,t+}^2(\mathbf{0})\right]. \tag{20}$$

The following corollary provides the investor's benefit from learning when there is one risky asset with two return predictors.

Corollary 2.3 *Under the condition of Theorem 2.2, the investor's benefit from each learning mode is given by:*

(i) *(Specialized learning) if $\theta_1\omega_1^2 e^{-\theta_1 K} > \theta_m b^2\omega_m^2$, then*

$$\begin{aligned}&\Delta(g_{1,t}, g_{m,t})\\ &= \frac{1}{2\gamma}\left(\frac{\left(a + b\phi_m g_{m,t} + \phi_1 g_{1,t}\right)^2}{\omega_1^2 + b^2\omega_m^2} + 1\right)\end{aligned}$$

$$\left(\frac{\omega_1^2 + b^2\omega_m^2}{\omega_1^2 e^{-\theta_1 K} + b^2\omega_m^2} - 1\right);$$

(ii) *(Diversified learning) if* $e^{-\theta_m K} \leqslant \frac{\theta_1\omega_1^2}{\theta_m b^2\omega_m^2} \leqslant e^{\theta_1 K}$, *then*

$$\Delta(g_{1,t}, g_{m,t}) = \frac{1}{2\gamma}\left(\frac{\left(a + b\phi_m g_{m,t} + \phi_1 g_{1,t}\right)^2}{\omega_1^2 + b^2\omega_m^2} + 1\right) \left[\frac{(\omega_1^2 + b^2\omega_m^2)e^{\frac{\theta_1\theta_m}{\theta_1+\theta_m}K}}{\omega_1^2\left(\frac{\theta_1\omega_1^2}{b^2\theta_m\omega_m^2}\right)^{-\frac{\theta_1}{\theta_1+\theta_m}} + b^2\omega_m^2\left(\frac{\theta_1\omega_1^2}{b^2\theta_m\omega_m^2}\right)^{\frac{\theta_m}{\theta_1+\theta_m}}} - 1\right];$$

(iii) *(Specialized learning) if* $\theta_1\omega_1^2 < \theta_m b^2\omega_m^2 e^{-\theta_m K}$, *then*

$$\Delta(g_{1,t}, g_{m,t}) = \frac{1}{2\gamma}\left(\frac{\left(a + b\phi_m g_{m,t} + \phi_1 g_{1,t}\right)^2}{\omega_1^2 + b^2\omega_m^2} + 1\right) \left(\frac{\omega_1^2 + b^2\omega_m^2}{\omega_1^2 + b^2\omega_m^2 e^{-\theta_m K}} - 1\right).$$

One can verify that Δ given in Corollary 2.3 is a continuous function of $K, \theta_1, \theta_m \in \mathbb{R}^+$. Then it is clear that the investor's benefit from learning is non-decreasing with the investor's attention capacity K and the learning efficiency parameters θ_1, θ_m. Specifically, increasing attention capacity leads to higher precision of all the information and hence enhances the benefit Δ in all three cases. Increasing θ_1 (resp. θ_m) improves the benefit from learning when the investor is not specialized in learning g_m (resp. g_1). Furthermore, the benefit from learning Δ is increasing in the squared mean of the asset return ($\left(a + b\phi_m g_{m,t} + \phi_1 g_{1,t}\right)^2$) and decreasing in the investor's risk aversion level (γ). These findings suggest that when the market is in extreme conditions, information acquisition and learning about the market are important for the effective management of portfolio risk, especially for aggressive investors.

THEOREM 2.4 *Suppose that there are two risky assets and the systematic component disappears. Therefore, there are only two firm-specific components* $g_{1,t}, g_{2,t}$. *Define the Sharpe ratio of the asset i as* $SR_{i,t} := \frac{a_i + \phi_i g_{i,t}}{\omega_i}$ *for i = 1, 2. Then the investor's optimal attention strategy is given by:*

(i) *(Specialized learning) if* $SR_{1,t}^2(e^{\theta_1 K} - 1) + e^{\theta_1 K} > SR_{2,t}^2(e^{\theta_2 K} - 1) + e^{\theta_2 K}$, *then* $\lambda_{1,t}^* = 1$ *and* $\lambda_{2,t}^* = 0$*;*
(ii) *(Specialized learning) if* $SR_{1,t}^2(e^{\theta_1 K} - 1) + e^{\theta_1 K} < SR_{2,t}^2(e^{\theta_2 K} - 1) + e^{\theta_2 K}$, *then* $\lambda_{1,t}^* = 0$ *and* $\lambda_{2,t}^* = 1$.

Moreover, the optimal investment strategy is

$$\boldsymbol{\pi}_{t+}^* = \frac{1}{\gamma}\begin{pmatrix} \frac{SR_{1,t}e^{\theta_1\lambda_{1,t}^* K}}{\omega_1} + \frac{1}{\omega_1^2}\left(e^{\theta_1\lambda_{1,t}^* K} - 1\right)\left(s_{1,t+} - \phi_1 g_{1,t}\right) \\ \frac{SR_{2,t}e^{\theta_2\lambda_{2,t}^* K}}{\omega_2} + \frac{1}{\omega_2^2}\left(e^{\theta_2\lambda_{2,t}^* K} - 1\right)\left(s_{2,t+} - \phi_2 g_{2,t}\right) \end{pmatrix}.$$

Theorem 2.4 indicates that when the risky assets are independent, the investor always chooses to be a specialist and learns the predictor with a higher $SR_{j,t}^2(e^{\theta_j K} - 1) + e^{\theta_j K}, j = 1, 2$. This quantity contains two terms: the former can be understood as the improvement in squared Sharpe ratio when $\lambda_j = 1$, and the latter is $\exp(2I(g_{j,t+1}; s_{j,t+}|\lambda_{j,t} = 1))$, which is an increasing function of the information quantity of $s_{j,t+}$ when the investor specializes in it. In other words, the investor should be specialized in learning about the asset with a large squared prior Sharpe ratio and a high learning efficiency. The effect of learning efficiency becomes more important as the capacity K increases.

Moreover, by taking the average of the investment strategy over the signals $s_{1,t+}$ and $s_{2,t+}$, we obtain

$$\mathbb{E}_t\left[\boldsymbol{\pi}_{t+}^*\right] = \frac{1}{\gamma}\left(\frac{SR_{1,t}e^{\theta_1\lambda_{1,t}^* K}}{\omega_1} \quad \frac{SR_{2,t}e^{\theta_2\lambda_{2,t}^* K}}{\omega_2}\right)^\top.$$

Since the investor is always a specialist, we find that the expected positions of the asset selected by the attention strategy are magnified by the $e^{\theta_j K}$ of the corresponding return predictor while the positions of the other asset are unchanged. That is, learning about one of the predictors motivates the investor to trade more in the related asset. The direction of the trade depends on the sign of the Sharpe ratio (equivalently, the expected return). Indeed, the attention strategy leads to an underdiversified portfolio that has a better overall performance than the classical portfolio without information acquisition.

It is worth noting that the finding that there are always corner solutions in the Theorem 2.4 is similar to that in Van Nieuwerburgh and Veldkamp (2010). However, different from Van Nieuwerburgh and Veldkamp (2010) who recommended using all the capacity to learn about the asset with the highest squared prior Sharpe ratio, our model implies that the investor should also take the learning efficiency into account.

We employ parameters that are in similar scales to the actual estimates from the Vanguard Real Estate Index Fund (VNQ) and VPU, based on the empirical study in Section 3. We let $\mathbf{b} = \mathbf{0}$, $SR_{1,t} = 1$ and $SR_{2,t} = 2$ and illustrate the impact of changes in the learning efficiency of the components on the proportion of attention allocation in figure 3. Figure 3 demonstrates that when two assets are independent, the investor is more likely to specialize in learning about the second asset which has a higher squared prior Sharpe ratio. The investor shifts all her attention to the asset 1 only when the learning efficiency of asset 2 is sufficiently lower than the learning efficiency of asset 1.

COROLLARY 2.5 *Under the conditions of Theorem 2.4, the investor's benefit from learning is given by:*

(i) *(Specialized learning) if* $SR_{1,t}^2(e^{\theta_1 K} - 1) + e^{\theta_1 K} > SR_{2,t}^2(e^{\theta_2 K} - 1) + e^{\theta_2 K}$, *then*

$$\Delta(g_{1,t}, g_{2,t}) = \frac{1}{2\gamma}\left(e^{\theta_1 K} - 1\right)\left(SR_{1,t}^2 + 1\right);$$

(ii) *(Specialized learning) if* $SR_{1,t}^2(e^{\theta_1 K} - 1) + e^{\theta_1 K} < SR_{2,t}^2(e^{\theta_2 K} - 1) + e^{\theta_2 K}$, *then*

$$\Delta(g_{1,t}, g_{2,t}) = \frac{1}{2\gamma}\left(e^{\theta_2 K} - 1\right)\left(SR_{2,t}^2 + 1\right).$$

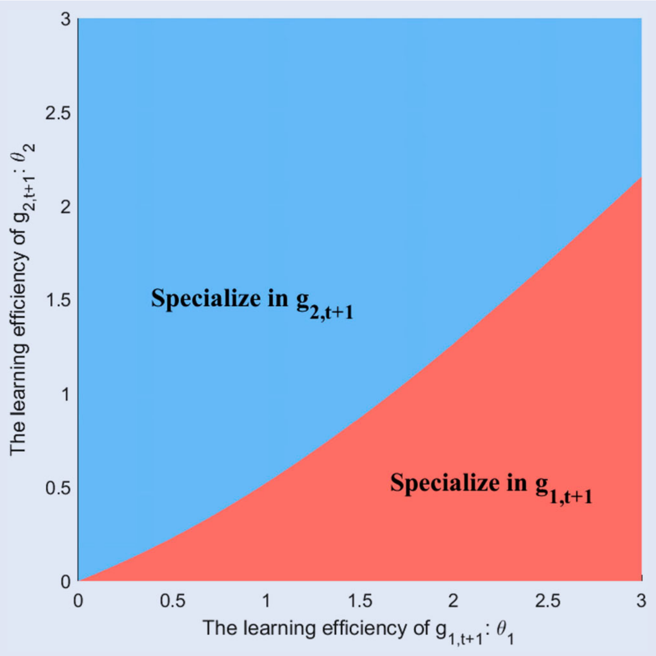


Figure 3. The optimal attention allocation in Theorem 2.4 with varying learning efficiency. We consider a market involving two risky assets, where we assume that the common systematic component is eliminated. The red domain represents that the investor focuses on learning the component $g_{1,t}$; the blue domain reflects that the investor specializes in the component $g_{2,t}$. Here, we use $SR_{1,t} = 1$, $SR_{2,t} = 2$ and $K = 1$.

As the results implied by Corollary 2.3, Theorem 2.4 and Corollary 2.5 indicate that the investor's benefit from learning increases with efficiency parameter θ_j when the investor is learning about the corresponding predictor g_j. The benefit from learning increases with attention capacity K and decreases with risk aversion γ no matter which predictor is learned. Moreover, when the investor is specialized in certain predictor g_j, the benefit from learning increases with the squared prior Sharpe ratio of the corresponding asset SR_j^2. This suggests that to earn more benefits, the investor's choice of information processing should prioritize assets with superior squared Sharpe ratios. Furthermore, as implied by Theorem 2.4, the attention strategy incurs portfolio under-diversification, that is, the weight of the asset in which the investor is specialized deviates from the classical Markowitz mean-variance portfolio. To have effective portfolio risk management, our strategy indicates that the investor should start with a well-diversified strategy according to Markowitz's methodology using prior information and then adjust it to concentrate on the asset/predictor according to the signal that is learned optimally.

3. Empirical study: a toy example

This section intends to numerically investigate interactions between attention strategy and investment strategy. To obtain a realistic parameter set in a simple manner, we use the empirical data of two assets in the market.

3.1. Data description

We look at two exchange-traded funds (ETFs), Vanguard Real Estate Index Fund (VNQ) and Vanguard Utilities Index Fund (VPU), as the two risky assets. VNQ measures the performance of the MSCI US Investable Market Real Estate 25/50 Index and invests in stocks issued by real estate investment trusts (REITs). While VPU tracks the returns of a benchmark index that measures the investment return of stocks in the utilities sector.

The price data of VNQ and VPU are downloaded from Yahoo Finance, and the data of market excess returns are obtained from Professor Kenneth R. French's online data library.† Our dataset includes monthly observations spanning from January 2007 to January 2022. We use the market excess returns, Mkt-RF, as the proxy for the systematic component $g_{m,t}$ and regress the historical returns of VNQ and VPU on Mkt-RF, respectively. The firm-specific components $g_{1,t}$ and $g_{2,t}$are then proxied by the residuals from these two regression models, where the excess returns of VNQ and VPU serve as the dependent variables. Here we consider that, consistent with the theoretical model, the investors can only trade two ETFs instead of $g_{1,t}, g_{2,t}, g_{m,t}$ separately.

Panel A of table 1 reports the estimation results from the regression models (1). The loading b_1 is more than twice as big as b_2, implying that VNQ returns are more sensitive to the systematic component than VPU returns. The R^2 in the regressions of VNQ and VPU returns on Mkt-RF are 0.5220 and 0.2422, respectively. These values suggest that the residual series of VPU captures a greater proportion of firm-specific information in comparison to the residual series of VNQ. As a result, when possessing the same level of knowledge about the market return, the investor is more uncertain about the return of VPU.

Furthermore, we also estimate the AR(1) processes of time series $\{g_{j,t}\}_{t\geqslant 0}$ for $j \in \{1, 2, m\}$, as shown in Panel B of table 1. The mean-reversion parameter ϕ_m for the systematic component is positive at 0.0963, while the mean-reversion parameters for the firm-specific components g_1 and g_2 are both negative, at -0.1085 and -0.1137 respectively. As a result, as expected, the systemic component exhibits greater persistence compared to the two firm-specific components. The conditional variances associated with these three components are relatively similar in magnitude, with the uncertainties of $g_{1,t+1}$ and $g_{m,t+1}$ being approximately twice as large as that of $g_{2,t+1}$. Moreover, we set the attention capacity as $K = 1$, and the risk-aversion coefficient as $\gamma = 1$.

Given the public availability of information regarding the systematic component, it is reasonable to assume that the investor possesses a higher level of efficiency in extracting information related to the systematic component compared to the firm-specific components (Peng and Xiong 2006). We select realistic parameters of learning efficiency: $\theta_1 = 0.6$, $\theta_2 = 0.8$, and $\theta_m = 1.6$. Indeed, in this particular setup, allocating all of the investor's attention to learning about g_m can lead to a variance reduction of the systematic component by approximately 80%, as compared to its prior variance. When the investor specializes in learning about either g_1 or g_2, the

† See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 1. The results of regression models and AR(1) processes for VNQ, VPU, and the stock market returns.

Panel A: Regression Results of VNQ and VPU			
	a_i	b_i	R^2
VNQ	– 0.0014	1.0673	0.5220
	(0.0036)	(0.0769)	
VPU	0.0036	0.4447	0.2422
	(0.0028)	(0.0592)	
Panel B: Estimation Results of AR(1) Processes			
	The residual of VNQ ($\{g_{1,t}\}_{t\geqslant 0}$)	The residual of VPU ($\{g_{2,t}\}_{t\geqslant 0}$)	Mkt-RF ($\{g_{m,t}\}_{t\geqslant 0}$)
ϕ_j	– 0.10848	– 0.1137	0.096273
	(0.0563)	(0.0776)	(0.0539)
ω_j^2	0.0021653	0.0012821	0.0021532
	(0.0002)	(0.0002)	(0.0002)

Notes: Panel A reports the regression results of excess returns of VNQ and VPU on the return predictors (equation (1)). Panel B presents the estimation results of the AR(1) processes (equation (2)) for the residuals of VNQ, VPU, and the stock market excess returns (Mkt-RF). Standard errors are reported in parentheses.

posterior variance of the corresponding firm-specific component is diminished by approximately 45% or 55%. Setting $\theta_1 < \theta_2$ makes the investor focus on predictors g_2, g_m and therefore simplifying the discussions. A more comprehensive sensitivity analysis of the impacts of θ_js will be conducted in Section 4.

3.2. *Impact of return predictors*

In this subsection, we examine the impact of the return predictors on optimal attention allocation and investment strategies. In the following, we will first focus on the attention and investment strategy in the second factor g_2 and later show all the strategies using heatmaps.

In figure 4(a,c), we set the systematic component $g_{m,t} = 0$, while in figure 4(b), we fix the specific component $g_{2,t} = 0$. In the three subfigures, we consider $g_{1,t}$ taking three possible values for $-0.1, 0$, and 0.1. Figure 4(a) demonstrates clear U-shaped relationships between the optimal attention strategy $\lambda^*_{2,t}$ and $g_{2,t}$, indicating that significant deviations of the return predictor from its long-term mean motivate the investor to pay more attention to that predictor. This is consistent with the findings in Andrei and Hasler (2020). Figure 4(c) present the expected investment strategy in asset 2, $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$, where we take the average to filter the fluctuations of the signals. The expected positions in asset 2 increase as $g_{2,t}$ deviates from 0. By comparing figure 4(a,c), we observe the interactions between the attention strategy and investment strategy. When the expected positions in asset 2 increase in absolute value, the investor allocates more attention to the related specific component $g_{2,t}$. Conversely, a larger $\lambda^*_{2,t}$ makes the investor more confident about the next-period component $g_{2,t+1}$ and the investor trades more aggressively in asset 2. This contributes to an underdiversified portfolio.

In figure 4(b), we let $g_{2,t} = 0$ and study the relation between $\lambda^*_{2,t}$ and $g_{m,t}$. We observe a novel inverted U-shaped relationship between the proportion of the investor's attention allocated to $g_{2,t}$ and the changes in $g_{m,t}$. This implies competition among the return predictors to attract the investor's attention. As $g_{m,t}$ takes more extreme values, the investor becomes increasingly attracted to it and in the meanwhile becomes less attentive to $g_{2,t}$ due to the attention capacity constraint.

Figure 5 presents a comprehensive result using heatmaps about how the optimal attention and expected investment vary with the return predictors $g_{1,t}$, $g_{2,t}$ and $g_{m,t}$. We choose $g_{m,t} \in \{-0.2, 0, 0.2\}$ and obtain three subfigures. In each subfigure, the first row displays the investor's optimal attention strategies $\lambda^*_{1,t}$, $\lambda^*_{2,t}$ and $\lambda^*_{m,t}$; the second row represents the expected investment strategy in three predictors, i.e. $\mathbb{E}_t\left[\pi^*_{1,t+}\right], \mathbb{E}_t\left[\pi^*_{2,t+}\right]$ and $b_1\mathbb{E}_t\left[\pi^*_{1,t+}\right] + b_2\mathbb{E}_t\left[\pi^*_{2,t+}\right]$.

In figure 5(b), the optimal attention strategies in the first row reveal that the investor allocates attention between $g_{2,t+1}$ and $g_{m,t+1}$, but consistently abstains from learning about $g_{1,t+1}$. The attention allocation strategy for $g_{2,t+1}$ shows that the investor tends to specialize in learning about the specific predictor $g_{2,t+1}$ when both $g_{1,t}$ and $g_{2,t}$ deviate substantially from their respective long-term means and have opposite signs (northwestern and southeastern corners). The attention allocated to the systematic predictor in figure 5(b) increases as $g_{1,t}$ and $g_{2,t}$ are close to each other. In other words, the investor's learning patterns change between specialized learning and diversified learning, depending on the magnitudes of the return predictors.

Looking solely at the attention strategy, it's not immediately clear why attention allocation should be affected by the signs of $g_{1,t}$ and $g_{2,t}$. However, this connection becomes evident when we examine the investment strategy. As indicated by the expected investment shares, the investor trades $g_{1,t+1}$ and $g_{2,t+1}$ aggressively for alpha payoffs when these predictors take extreme values and have different signs (northwestern and southwestern corners) while adopts a relatively

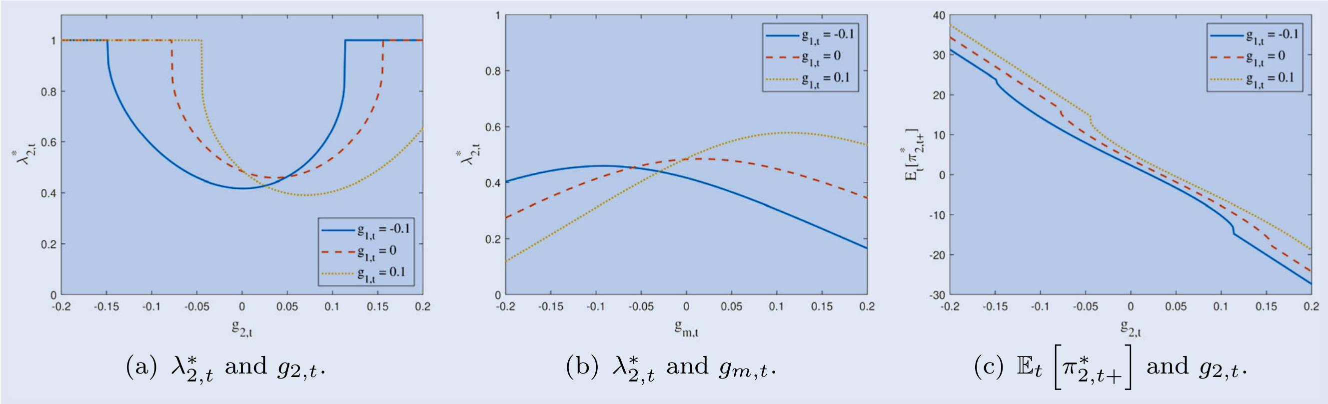


Figure 4. Response of the optimal attention strategy and the expected investment strategy to changes in the return predictors. (a) $\lambda^*_{2,t}$ and $g_{2,t}$. (b) $\lambda^*_{2,t}$ and $g_{m,t}$ and (c) $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ and $g_{2,t}$. We plot the optimal attention strategy $\lambda^*_{2,t}$ with varying firm-specific component $g_{2,t}$ and the systematic component $g_{m,t}$. We also plot the expected investment strategy $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ with varying $g_{2,t}$. In (a) and (c), we fix the systematic component $g_{m,t} = 0$, while in (b), we consider $g_{2,t} = 0$. The attention strategy $\lambda^*_{1,t}$ is always equal to 0 and therefore $\lambda^*_{m,t} = 1 - \lambda^*_{2,t}$. To focus on the strategies for the second specific predictor, the other components of the optimal strategies are not reported.

conservative strategy in other domains. Combining the first and the second rows in figure 5(b), we observe a significant correspondence between the attention allocated and investment strategy. The investor intends to adopt an aggressive long-short alpha strategy, and the large positions in the second asset, in turn, motivate the investor to learn more about $g_{2,t+1}$. On the other hand, we observe discontinuous changes in expected positions of $g_{2,t+1}$ and $g_{m,t+1}$, which implies that the improved precision of predicting $g_{2,t+1}$ further strengthens the alpha strategy.

In figure 5(a,c), where the systematic component $g_{m,t} = -0.2$ and $g_{m,t} = 0.2$ respectively, it is clear from the third column that the investor becomes more likely to long or short large shares of $g_{m,t+1}$ on average over the signals. And correspondingly, the attention strategy $\lambda^*_{m,t}$ increases in the domain where the expected investment in $g_{m,t+1}$ is aggressive. In these situations, the investor utilizes the market trend, i.e. the mean-reverting pattern of $g_{m,t}$, to earn benefits. This beta strategy incurs more positions in $g_{1,t+1}$ and therefore motivates the investor to diversify attention and learn about $g_{1,t+1}$ sometimes.

Figure 6 outlines the interactions between the investor's attention allocation and portfolio strategy. The prior information can indicate possible investment opportunities. Motivated by this, the investor starts by closely observing the signals about the corresponding predictors. The incrementally increased attention to these signals resolves the portfolio uncertainty and contributes to the investor's growing confidence in future payoffs. Consequently, the investor constructs an underdiversified portfolio based on the posterior information.

In summary, we find the attention strategy tends to match up with the investment strategy in the sense that it reduces the uncertainty of those predictors significant in the portfolio, and in turn, improves the absolute positions in the predictors. The investor may exhibit different investment skills including picking alpha and betting on beta depending on the deviations of predictor from their long-term means.

Furthermore, we illustrate how the investor's benefit from learning, $\Delta(g_{1,t}, g_{2,t}, g_{m,t})$, varies across different values of return predictors in figure 7. In general, the benefit from learning is significant when the three predictors take extreme values, and the benefit from learning is mild when they are around zero, consistent with the theoretical predictions in specific cases in Corollaries 2.3 and 2.5. Combining figure 7 and the plots investment strategy of figure 5, we observe that the domains of large benefit from learning are mostly consistent with the domains when the investor takes aggressive long-short trading strategies in two assets. This suggests that, in practice, especially during periods of significant market fluctuation, for investors trading aggressively such as hedge funds, allocating attention effectively is crucial for proper risk management.

3.3. *Impact of mean-reversion parameter*

The coefficient of the return predictor, i.e. ϕ, also plays a crucial role in affecting the investor's decision-making process. When $\phi_j = 0$, $g_{j,t+1} = \epsilon_{j,t+1}$ and this makes $g_{j,t+1}$ unpredictable at time t. We consider $|\phi_j| < 1$ so that $g_{j,t}$ is a stationary process. In this case, a large ϕ_j means the predictor has a slow mean-reverting speed.

To keep the discussion concise, we only study the relations between the investor's optimal strategy $\lambda^*_{2,t}$, $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ and the mean-reverting speed ϕ_2. More discussions on parameter sensitivity will be given in Section 4. In figure 8(a,c), we fix $g_{m,t} = 0$ and choose $g_{2,t} \in \{-0.1, 0, 0.1\}$. When $g_{2,t} = 0$, both $\lambda^*_{2,t}$ and $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ are unchanged since $g_{2,t+1}$ is purely a noise term independent of ϕ_2 in this case. When $g_{2,t} \neq 0$, we observe a U-shaped relationship between $\lambda^*_{2,t}$ and ϕ_2. Indeed, when $|\phi_2|$ is large, no matter $g_{2,t}$ is fast mean-reverting or slow mean-reverting, $|g_{2,t+1}|$ is more likely to take extreme values. This motivates the investor to assign a higher level of attention to learning about $g_{2,t+1}$. Figure 8(c) also demonstrates that the investor tends to hold more $g_{2,t+1}$ in the portfolio when ϕ_2 takes extreme values. In other words, $\lambda^*_{2,t}$ also matches with $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ in this case.

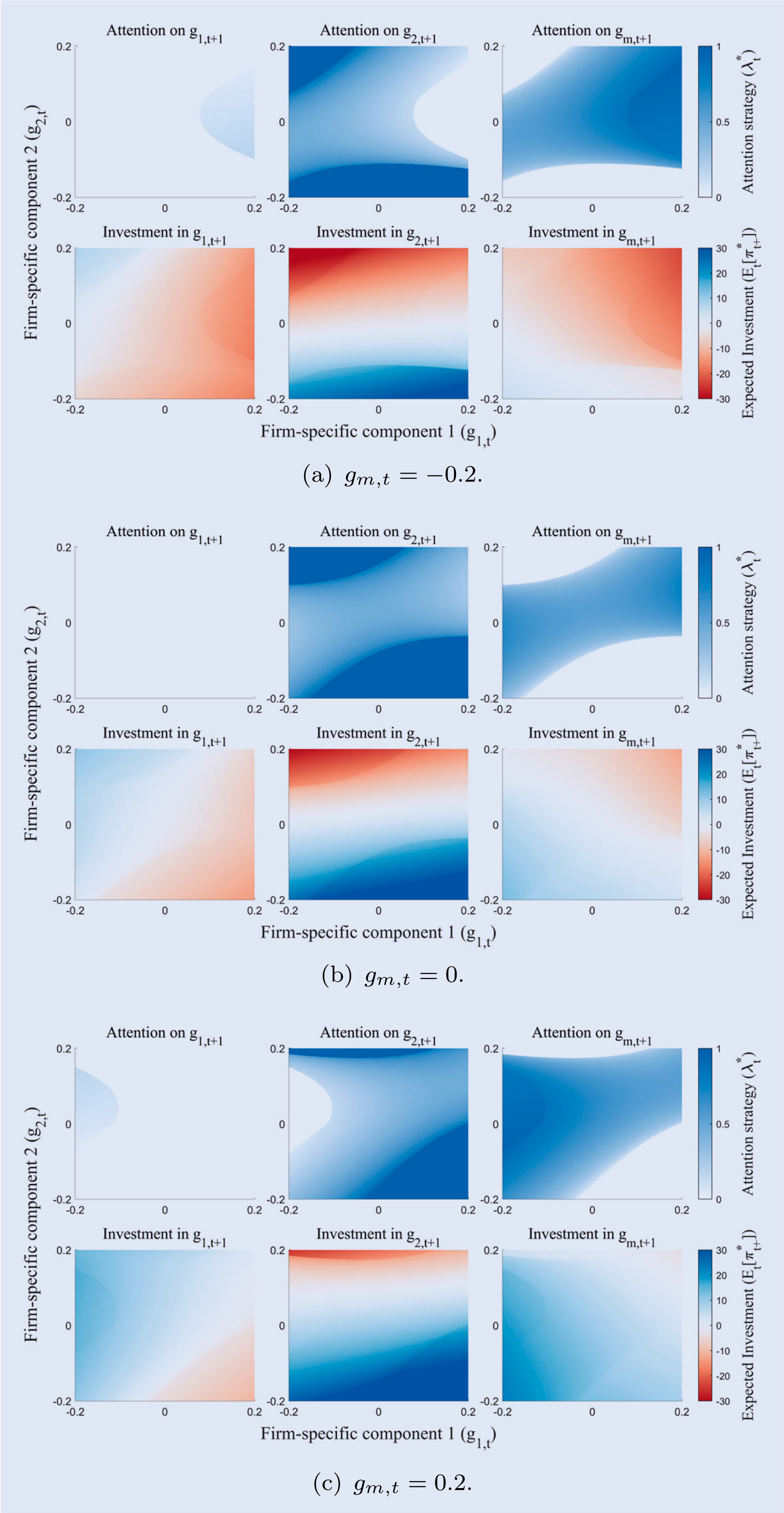


Figure 5. The optimal attention strategies and the expected investment shares for the single-period investment in the toy example. (a) $g_{m,t} = -0.2$. (b) $g_{m,t} = 0$ and (c) $g_{m,t} = 0.2$. We choose the systematic component $g_{m,t}$ to be -0.2 (a), its long-term mean 0 (b) and 0.2 (c), respectively. The color gradient in the first row corresponds to changes in the proportion of attention allocation, and the color transition from white to the darkest blue corresponds to a range of proportions from 0 to 1. The color gradient in the second row reflects changes in expected investment in three predictors. The color transition from the darkest red to the darkest blue corresponds to the expected investment in predictors, ranging from deep short to deep long positions.

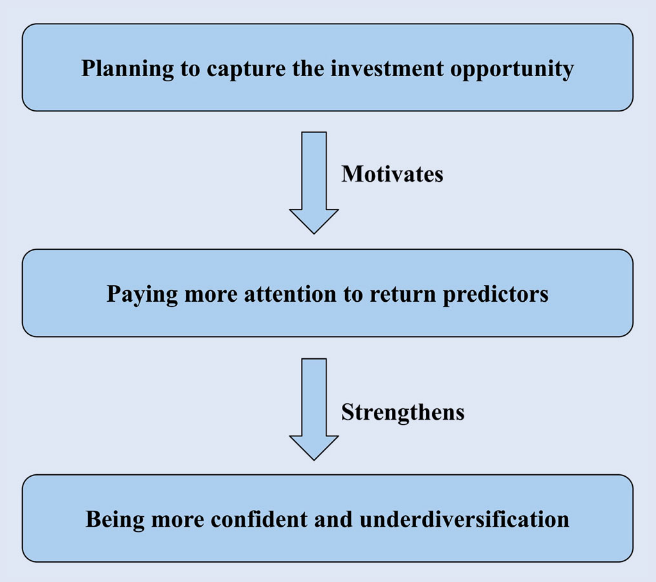


Figure 6. The interactions between the investor's attention allocation and portfolio strategy.

The inverted U-shaped pattern featured in figure 8(b) further demonstrates the competition among the return predictors, where we choose $g_{m,t} = 0.1$ and plot $\lambda^*_{2,t}$ with varying ϕ_m. As $|\phi_m|$ increases, the investor expects to have a more accurate estimation of $g_{m,t+1}$, which is more likely to take extreme values. Consequently, the investor attention is attracted by $g_{m,t+1}$, and less attention is paid to $g_{2,t+1}$.

4. Comparative statics analysis

The toy example presented above indicates a U-shaped relationship between the optimal allocation of attention and the corresponding return predictor. The new implication of our multi-asset framework is the competition among information sources for the investor's limited attention, which is an inverted U-shaped relationship between the allocation of attention by predictors and other information sources. Whether the investor should pursue aggressive trading to exploit alpha opportunities or take a conservative portfolio

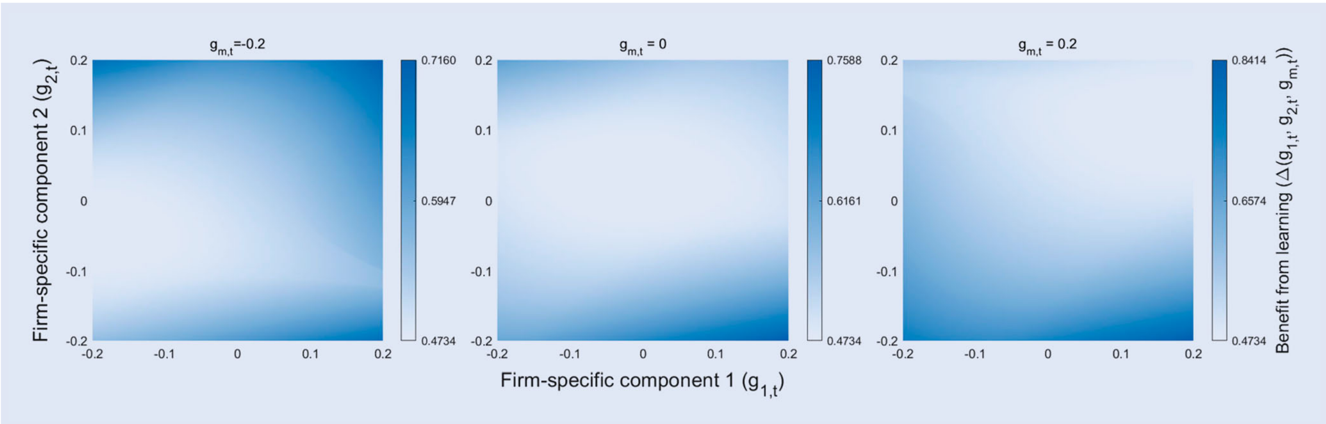
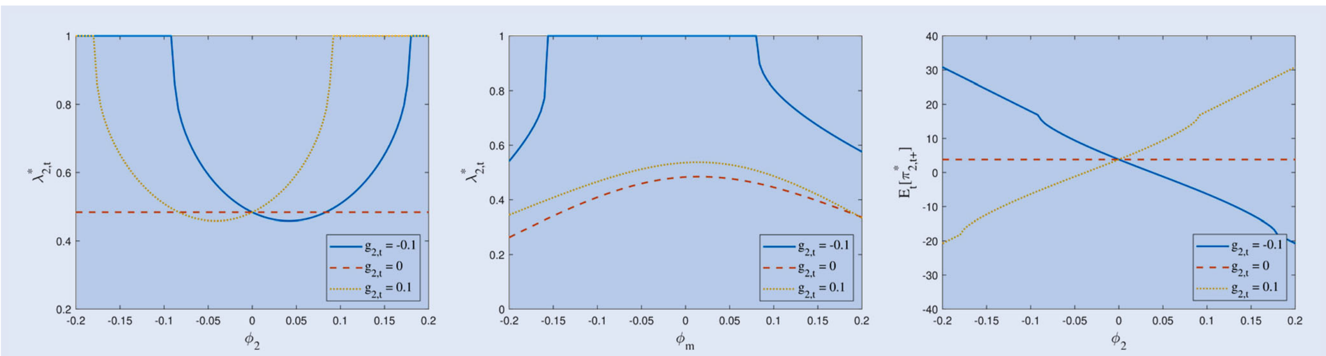
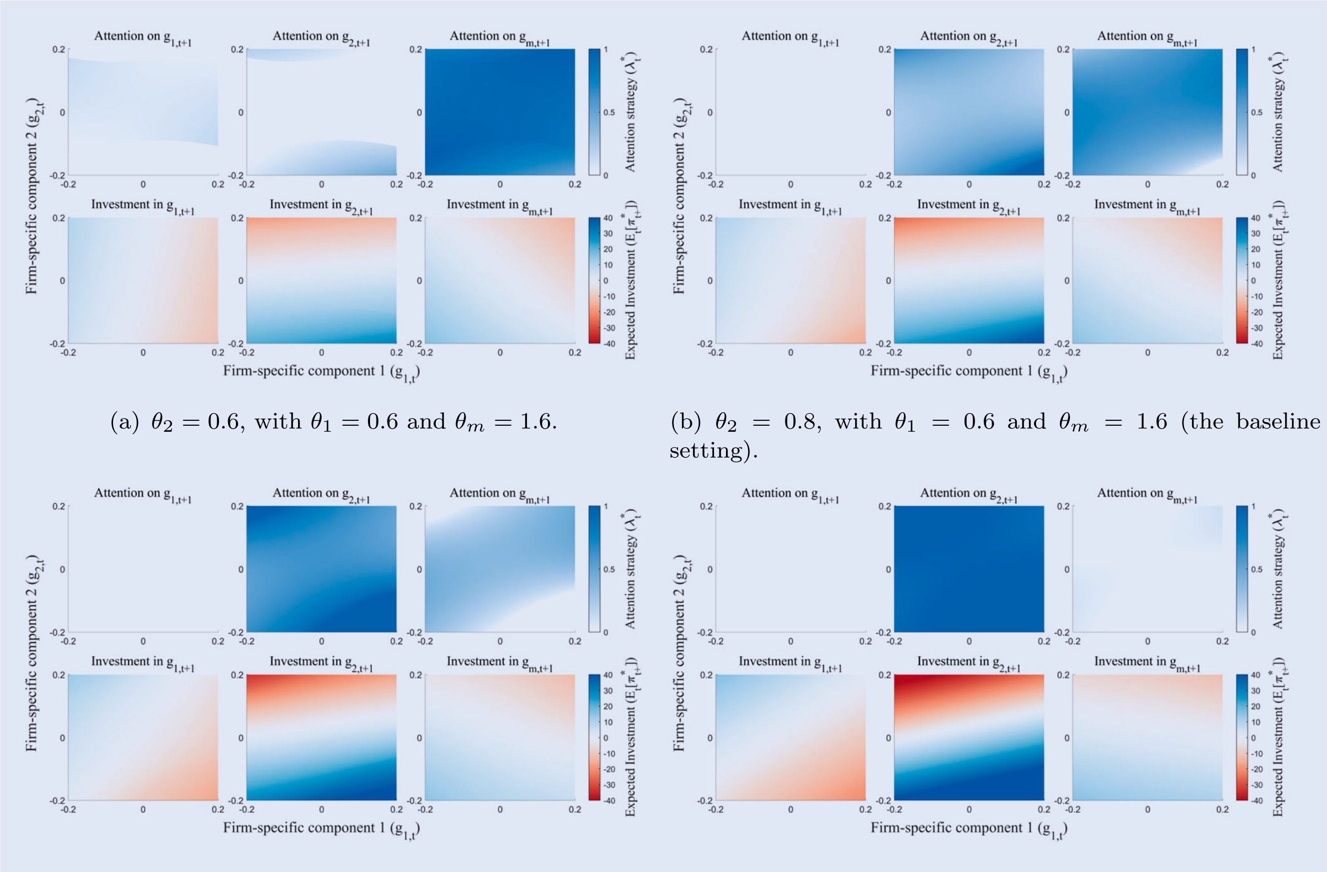


Figure 7. The investor's benefits from learning for the single-period investment in the toy example. We choose the systematic component $g_{m,t}$ to be -0.2, its long-term mean 0 and 0.2, respectively. The color gradient corresponds to changes in the benefit from learning.



(a) $\lambda^*_{2,t}$ and ϕ_2. (b) $\lambda^*_{2,t}$ and ϕ_m. (c) $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ and ϕ_2.

Figure 8. Response of the optimal attention strategy and the expected investment strategy to changes in mean-reversion parameters. (a) $\lambda^*_{2,t}$ and ϕ_2. (b) $\lambda^*_{2,t}$ and ϕ_m and (c) $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ and ϕ_2. We plot the optimal attention strategy $\lambda^*_{2,t}$ varying with the mean-reversion parameters ϕ_2 and ϕ_m, as well as the expected investment strategy $\mathbb{E}_t\left[\pi^*_{2,t+}\right]$ varying with ϕ_2 for $g_{2,t} \in \{-0.1, 0, 0.1\}$ in the above figures. In (a,c), we choose $g_{1,t} = g_{m,t} = 0$, while in (b), we consider $g_{1,t} = 0$ and $g_{m,t} = 0.1$. The attention strategy $\lambda^*_{1,t}$ is always equal to 0 and thus is not reported.



(a) $\theta_2 = 0.6$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$.

(b) $\theta_2 = 0.8$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$ (the baseline setting).

(c) $\theta_2 = 1$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$.

(d) $\theta_2 = 1.6$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$.

Figure 9. The optimal attention allocation strategy and the expected investment strategy (averaged over signals) varying with the learning efficiency θ_2. (a) $\theta_2 = 0.6$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$. (b) $\theta_2 = 0.8$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$ (the baseline setting). (c) $\theta_2 = 1$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$ and (d) $\theta_2 = 1.6$, with $\theta_1 = 0.6$ and $\theta_m = 1.6$. The values of the remaining model parameters are set to those in the baseline setting (21). Moreover, we fix the systematic predictor $g_{m,t} = 0$.

position for beta investment hinges on the extent to which various return predictors deviate from their respective long-term means. This phenomenon highlights the dynamics of attention allocation and portfolio rebalancing in our model. For simplicity, we focus our sensitivity analyzes on parameters related to the firm-specific predictor $g_{2,t}$. The obtained conclusions also apply to $g_{1,t}$.

We include two risky assets in the numerical simulations following the setup of the toy example. In alignment with the estimation results from the toy example, we choose the following baseline setting:

$$\begin{aligned} &\boldsymbol{a} = (-0.0015, 0.0035)^\top, \quad \boldsymbol{b} = (1, 0.5)^\top, \\ &\boldsymbol{\theta} = (0.6, 0.8, 1.6)^\top, \\ &\boldsymbol{\phi} = (-0.1, -0.1, 0.1)^\top, \quad \boldsymbol{\omega}^2 = (0.002, 0.001, 0.002)^\top, \\ &K = 1, \ \gamma = 1. \end{aligned} \tag{21}$$

We set $g_{m,t} = 0$ in the analyzes of all the parameters except for the analysis of b_2 where we consider $g_{m,t} = 0.1$. When we vary one parameter, all the remaining parameters are kept fixed and are equal to those in the baseline setting.

4.1. *Sensitivity analysis of optimal strategy on learning efficiency*

We first examine the impact of the learning efficiency θ_j on the investor's optimal attention allocation ratios and the expected investment shares. Figure 9 shows the results. As expected, when the learning efficiency θ_2 increases, the investor tends to allocate more attention to learning about $g_{2,t+1}$ because it is more efficient to improve the portfolio performance in this way. In the meanwhile, increasing θ_2 also makes learning about other predictors less attractive and the corresponding attention decreases. It is worth noting that when θ_2 is equal to θ_m, namely 1.6, the investor chooses to be almost specialized in learning about the firm-specific predictor $g_{2,t+1}$ and is no longer attentive to the public predictor $g_{m,t+1}$. In this case, concentrating all attention on resolving the uncertainty of idiosyncratic predictors has greater efficiency in maximizing portfolio performance. As θ_2 increases, the expected investment in $g_{2,t+1}$ becomes more extreme whereas the positions in other predictors are less changed. This finding suggests that the improvements in learning efficiency make the investor more aggressive in trading the corresponding return components and this further leads to a more underdiversified portfolio.

4.2. Sensitivity analysis of optimal strategy on risk exposure

Through our theoretical analysis, it becomes evident that the risk exposure of excess return plays a pivotal role in influencing the investor's decision-making process. We consider multiple values for b_2, and the results are presented in figure 10.

When b_2 is positive and increases, indicating the amount of systematic component increases in the return of asset 2, the investor shifts attention from the specific predictor $g_{2,t+1}$ towards the systematic predictor $g_{m,t+1}$. Initially, the investor is more likely to focus on $g_{2,t+1}$, then sometimes spreads out attention, and finally, specializes in $g_{m,t+1}$ most of the time. In the meanwhile, as b_2 increases, the expected investment in $g_{2,t+1}$ decreases, and the expected investment in $g_{m,t+1}$ increases. This suggests the investor transitions from an active asset-picking strategy to a passive market-trend investment strategy as the loading coefficient increases.

When the systematic component negatively predicts the return of asset 2, i.e. $b_2 < 0$, the investor's optimal attention allocation and investment strategies show similar patterns as the cases when $b_2 > 0$. However, there is a distinction between cases where b_2 has different signs. Specifically, when $b_2 < 0$, we observe specialized learning on $g_{2,t+1}$ in the northeastern and southwestern corners and diversified learning in the other two corners. These learning patterns' locations differ from those observed when $b_2 > 0$. We note that when $b_2 > 0$, an alpha investment strategy that hedges the systematic component lies on the northeastern and southwestern corners; when $b_2 < 0$, such an alpha strategy lies in the northwestern and southeastern corners. Therefore, the distinct signs of b_2 alter the investment patterns, consequently motivating varied learning patterns.

4.3. Sensitivity analysis of optimal strategy on mean-reversion parameter

In figure 11, we further examine the impacts of ϕ_2 on the investor's attention allocation and portfolio choice. First, when $\phi_2 = 0$ (see figure 11(c)), $g_{2,t+1}$ can be regarded as white noise and is unpredictable. The investor chooses diversified learning with more attention paid to $g_{m,t+1}$. In this case, the attention strategy is independent of the magnitudes of return predictors. In other cases where $\phi_2 \neq 0$, both $g_{2,t}$ and $g_{m,t}$ are useful in predicting $g_{2,t+1}$ and $g_{m,t+1}$. The competition of two predictors incurs different learning patterns.

Specifically, as ϕ_2 deviates from 0, meaning that the process g_2 is fast mean-reverting or slow mean-reverting, the investor's attention may be diverted from $g_{m,t+1}$ to $g_{2,t+1}$ and may even become concentrated on $g_{2,t+1}$. As shown in figure 11, when $\phi_2 < 0$, $\lambda^*_{2,t} = 1$ appears in the northwestern and southeastern corners; when $\phi_2 > 0$, $\lambda^*_{2,t} = 1$ appears in the northwestern and southwestern corners. In these regions, the investor constructs a long-short portfolio using asset 1 and asset 2. The significant positions of $g_{2,t+1}$ make the investor pay more attention to it. Indeed, the increase of $|\phi_2|$ can be explained as an improvement in the predictability of $g_{2,t}$ for $g_{2,t+1}$. Therefore, figure 11 demonstrates that to reduce portfolio uncertainty, more extreme mean-reverting speeds render current information more valuable and will motivate the investor to learn about the corresponding predictors, especially when these predictors are significant in the positions. We also note that in all the scenarios, the investor always ignores the signals about $g_{1,t+1}$ due to the low efficiency of learning about $g_{1,t+1}$.

4.4. Sensitivity analysis of optimal strategy on attention capacity

We explore the impact of an investor's attention capacity on her attention allocation and portfolio selection by varying the values of parameter K. Figure 12 shows the results.

In the baseline setting, the investor's information capacity is set to 1. In this scenario, the investor demonstrates a noticeably greater interest in movements of both the systematic component $g_{m,t+1}$ and the firm-specific component $g_{2,t+1}$. This diversified learning approach leads to profitable outcomes. When the capacity is severely limited ($K = 0.5$), there is a notable expansion in the regions where the investor specializes. Specifically, the investor focuses on signals about $g_{m,t+1}$ in most cases, and shifts focus to the information about $g_{2,t+1}$ when taking a substantial long position in $g_{2,t+1}$. Furthermore, the investor constructs a more conservative portfolio when the information capacity is more restrictive. Conversely, when the capacity is significantly larger, as illustrated in figure 12(c,d), the investor always chooses diversified learning, and the attention paid to $g_{m,t+1}$ decreases as K increases. The investor with sufficiently higher capacity ($K = 10$) allocates equal weights to both firm-specific components to extract more idiosyncratic signals and leverages alpha returns. In this case, the investor fully hedges the systematic component, resulting in an expected investment of zero in $g_{m,t+1}$. Consequently, there is no necessity to learn about $g_{m,t+1}$, and the public signals are disregarded.

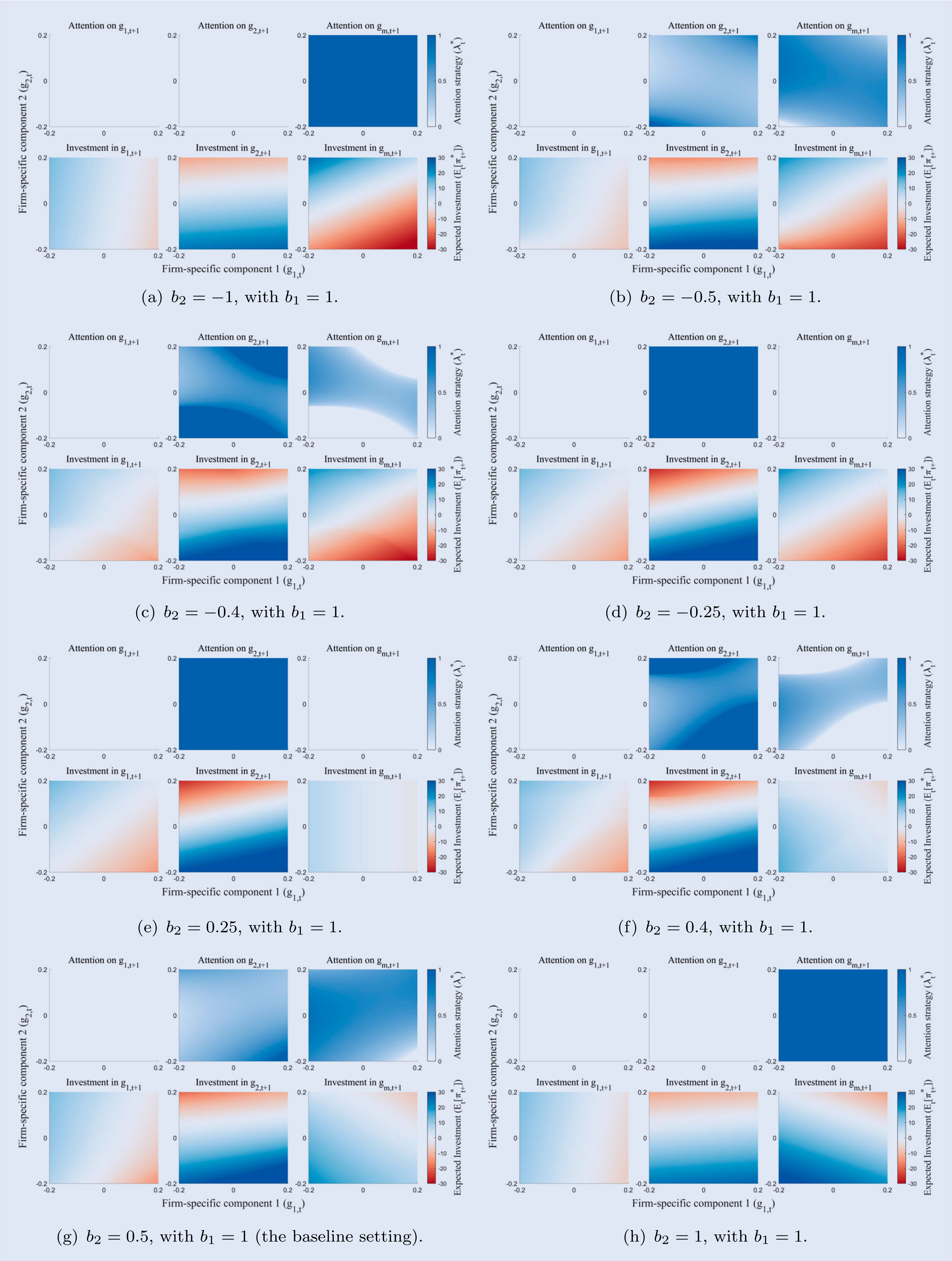
In summary, when an investor has very limited information capacity, they focus on understanding the overall market trends and aim for steady returns. However, with more information capacity, the investor shifts focus to unique signals from individual assets, taking more risks for potentially higher returns. The attention allocation and portfolio choice are closely related.

5. Extensions

This section involves extensions of our previous investigations in various aspects. We first apply the static optimization problem to the cryptocurrency market to show the economic interpretations of our models in different markets. We then extend our model to the dynamic attention allocation and portfolio selection problem with multiple assets over multiple periods. Finally, we study the impact of attention allocation on asset pricing by investigating the economic equilibrium of our model within the framework of CAPM theory.

5.1. Applications to the cryptocurrency market

In addition to the ETFs in the stock market discussed earlier, we extend our model to include the rapidly growing



(a) $b_2 = -1$, with $b_1 = 1$.

(b) $b_2 = -0.5$, with $b_1 = 1$.

(c) $b_2 = -0.4$, with $b_1 = 1$.

(d) $b_2 = -0.25$, with $b_1 = 1$.

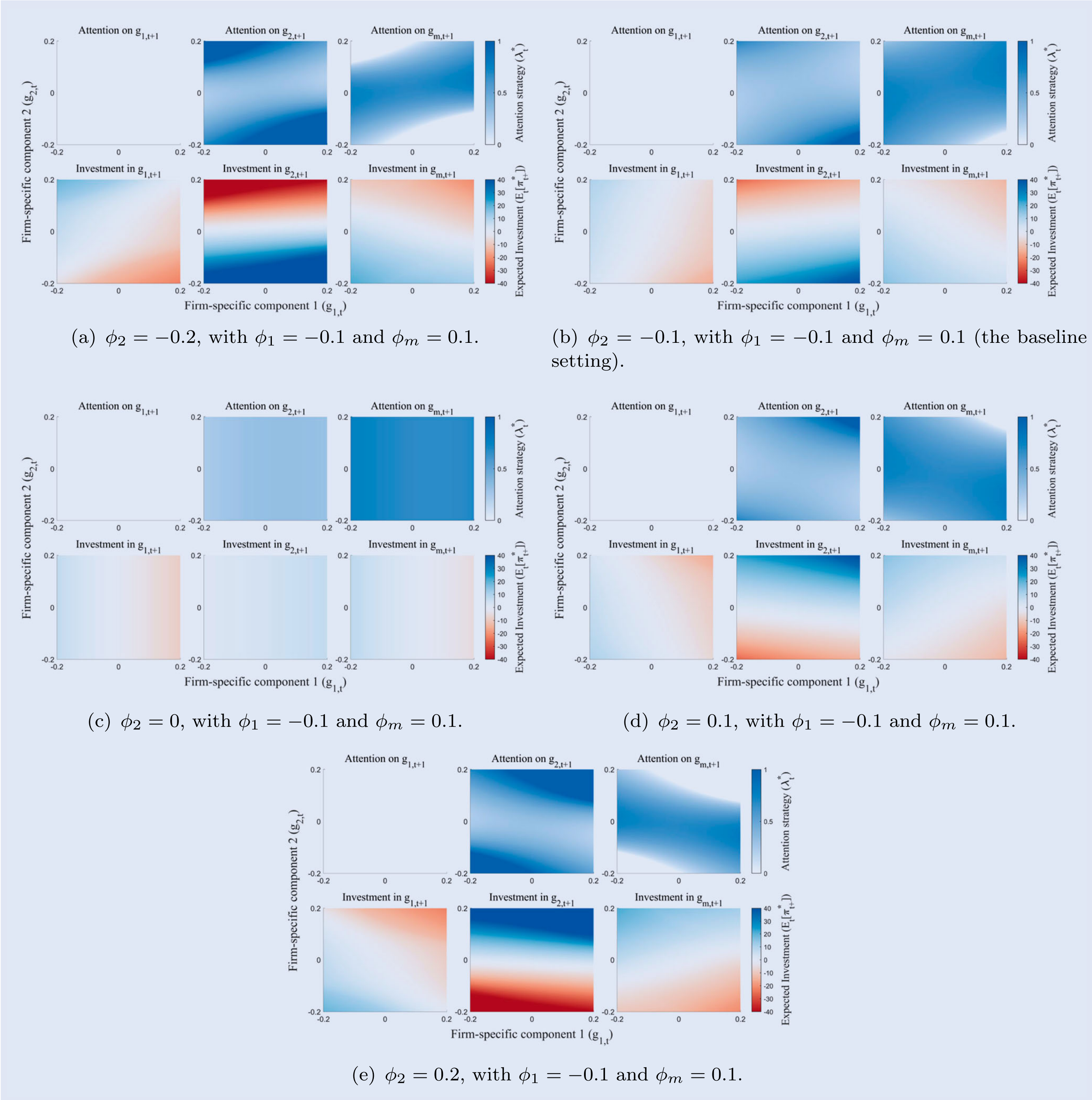
(e) $b_2 = 0.25$, with $b_1 = 1$.

(f) $b_2 = 0.4$, with $b_1 = 1$.

(g) $b_2 = 0.5$, with $b_1 = 1$ (the baseline setting).

(h) $b_2 = 1$, with $b_1 = 1$.

Figure 10. The optimal attention allocation strategy and the expected investment strategy (averaged over signals) varying with the risk exposure b_2. (a) $b_2 = -1$, with $b_1 = 1$. (b) $b_2 = -0.5$, with $b_1 = 1$. (c) $b_2 = -0.4$, with $b_1 = 1$. (d) $b_2 = -0.25$, with $b_1 = 1$. (e) $b_2 = 0.25$, with $b_1 = 1$. (f) $b_2 = 0.4$, with $b_1 = 1$. (g) $b_2 = 0.5$, with $b_1 = 1$ (the baseline setting) and (h) $b_2 = 1$, with $b_1 = 1$. The values of the remaining model parameters are set to those in the baseline setting (21). We choose the systematic component $g_{m,t} = 0.1$.



(a) $\phi_2 = -0.2$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$.

(b) $\phi_2 = -0.1$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$ (the baseline setting).

(c) $\phi_2 = 0$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$.

(d) $\phi_2 = 0.1$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$.

(e) $\phi_2 = 0.2$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$.

Figure 11. The optimal attention allocation strategy and the expected investment strategy (averaged over signals) varying with the mean reversion parameter ϕ_2. (a) $\phi_2 = -0.2$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$. (b) $\phi_2 = -0.1$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$ (the baseline setting). (c) $\phi_2 = 0$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$. (d) $\phi_2 = 0.1$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$ and (e) $\phi_2 = 0.2$, with $\phi_1 = -0.1$ and $\phi_m = 0.1$. The values of the remaining model parameters are set to those in the baseline setting (21). We choose the systematic component $g_{m,t} = 0$.

cryptocurrency market. While our primary analysis focuses on the well-established equity market and traditional asset classes, this extension explores whether our model's insights into the investor's attention allocation and portfolio choices apply to emerging markets. The cryptocurrency market, characterized by high volatility and a strong reliance on financial technology, not only attracts global investors but has also garnered extensive academic interest. Liu and Tsyvinski (2021) and Liu *et al.* (2022) demonstrate empirically that momentum and investor attention are important predictors of cryptocurrency market returns. Their findings confirm the theoretical predictions of Peng and Xiong (2006) and Hou *et al.* (2009) that limited attention contributes to the momentum effects among high-attention cryptocurrencies. Motivated by existing theoretical insights and empirical findings, we estimate the factor models presented in equation (1) using cryptocurrency-specific predictors and investigate how investor attention and portfolio allocations can be optimally determined in the cryptocurrency market.

As of July 30, 2024, Bitcoin (hereafter BTC) dominates the cryptocurrency market with a market share of 54.91%, followed by Ethereum (16.68%, hereafter ETH), Tether (4.78%), Solana (3.54%), BNB (3.49%), and others (16.59%). Given that the aggregated market capitalization of BTC and ETH

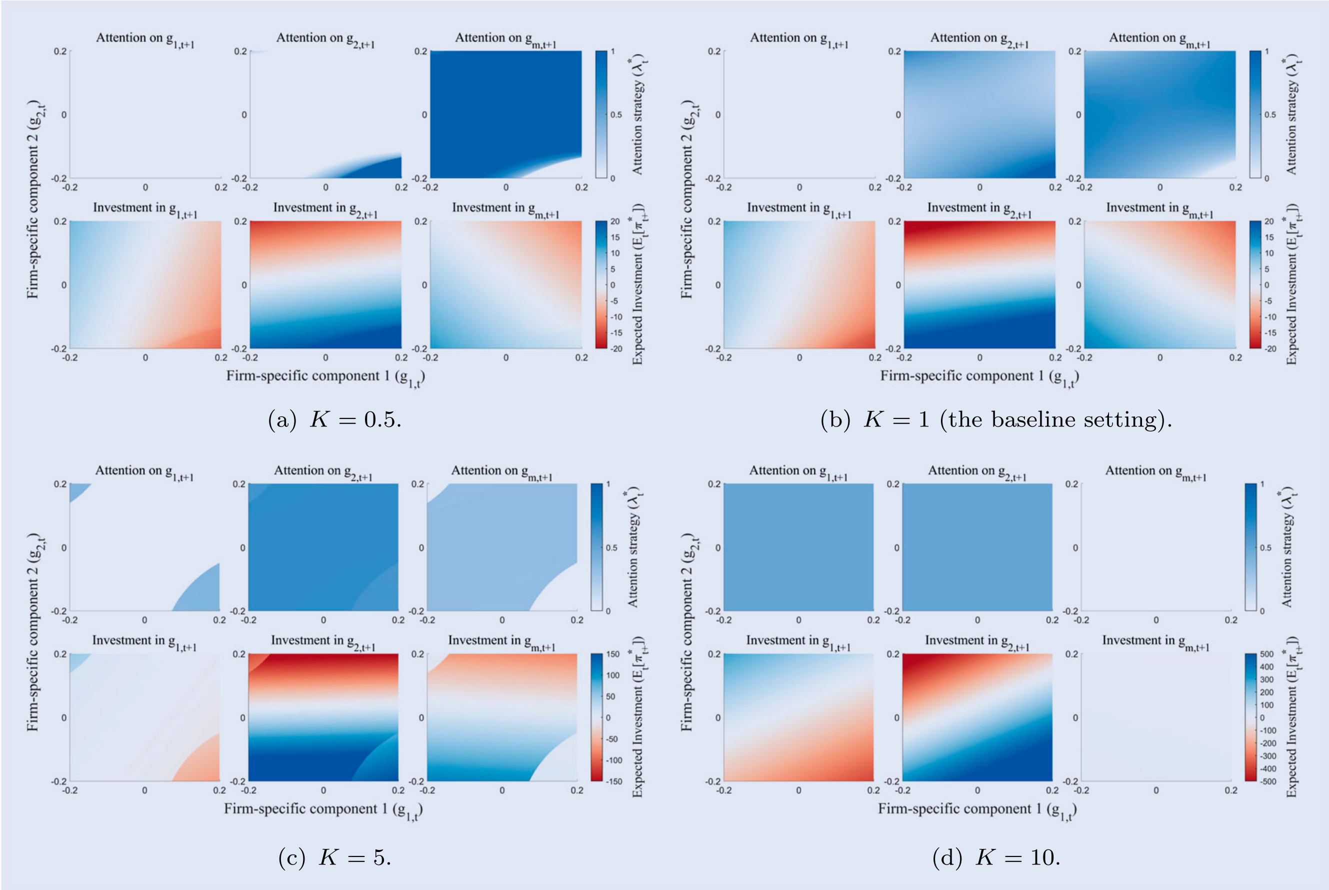


Figure 12. The optimal attention allocation strategy and the expected investment strategy (averaged over signals) varying with the information capacity K. (a) $K = 0.5$. (b) $K = 1$ (the baseline setting). (c) $K = 5$ and (d) $K = 10$. The values of the remaining model parameters are set to those in the baseline setting (21). We choose the systematic component $g_{m,t} = 0$.

accounts for over 70% of the total cryptocurrency market, our analysis focuses on these two largest and best-known cryptocurrencies. The excess returns of BTC and ETH are each regressed separately on the excess returns of the broader cryptocurrency market, $g_{m,t}$, which is calculated using the Crypto200 (hereafter CMC) Index as a proxy. The CMC200 Index tracks the performance of the top 200 cryptocurrencies by market capitalization, providing a comprehensive benchmark of market-wide excess returns.† The residuals $g_{1,t}$ and $g_{2,t}$ from the regression model (1) represent the return predictors for forecasting the future excess returns of BTC and ETH, respectively.

Panel A of table 2 shows that while the loadings b_1 and b_2 are both approximately 1, the CMC200 Index excess returns explain a larger proportion of the excess returns of BTC ($R^2 = 0.8897$) compared to ETH ($R^2 = 0.6932$). Notably, the R^2s for both BTC and ETH are higher than those of VNQ and VPU in the equity market, reflecting the strong correlation between the excess returns of the two largest cryptocurrencies and the coin market excess returns. Panel B of table 2 presents results similar to those in Panel B of table 1, as the systematic component exhibits greater persistence with positive ϕ_m compared to the cryptocurrency-specific components of BTC and ETH. The variances of these return predictors are extremely small, particularly for the return predictor of BTC, which is an order of magnitude smaller than the other two return predictors. Based on market capitalization and the estimated results, we set the learning efficiency coefficients to $\theta_1 = 1.6$, $\theta_2 = 1.4$, and $\theta_m = 2$, as the higher volatility and rapid fluctuations in the cryptocurrency market necessitate greater learning efficiency for the investor to track and respond effectively to market trends.

We show in figure 13 the investor's optimal attention allocation ratios and shareholdings in a portfolio consisting of BTC, ETH, and CMC200 Index in three scenarios $g_{m,t} = -0.4, 0$, and 0.4.‡ In figure 13, the optimal strategies suggest that when both the return predictors of BTC and ETH, $g_{1,t}$ and $g_{2,t}$, deviate substantially from zeros and have opposite signs (northwestern and southeastern corners), the investor chooses diversified learning in two cryptocurrency-specific components and uses long-short positions to pick alpha returns. Due to the higher information content for $g_{2,t+1}$ (suggested by the lower R^2 in Panel A of table 2), the investor has more incentive to process information about ETH relative to

† The weekly price data for BTC, ETH, and CMC200 Index are downloaded from Yahoo Finance during the period from August 18, 2020 to July 30, 2024. The market capitalization data are sourced from coinmarketcap.com and are reported in U.S. dollars.

‡ According to the estimated results, the range of values for return predictors is notably broader in the cryptocurrency market compared to that observed in the previous equity market analysis, where the bounds were set to ±0.2. Thus, we set the upper and lower bounds of the return predictors to ±0.4 in this subsection.

Table 2. The results of regression models and AR(1) processes for BTC, ETH, and CMC200 Index.

	Panel A: Regression Results of BTC and ETH		
	a_i	b_i	R^2
BTC	7.47×10^{-4}	0.9563	0.8897
	(0.0021)	(0.0237)	
ETH	0.0034	1.0368	0.6932
	(0.0043)	(0.0485)	
	Panel B: Estimation Results of AR(1) Processes		
	The residual of BTC ($\{g_{1,t}\}_{t\geqslant 0}$)	The residual of ETH ($\{g_{2,t}\}_{t\geqslant 0}$)	CMC200 Index excess returns ($\{g_{m,t}\}_{t\geqslant 0}$)
ϕ_j	– 0.19679	– 0.10528	0.10258
	(0.040030)	(0.045074)	(0.081316)
ω_j^2	0.000850	0.003668	0.007854
	(0.000055)	(0.000161)	(0.000619)

Notes: Panel A reports the regression results of excess returns of BTC and ETH on the return predictors (equation (1)). Panel B presents the estimation results of the AR(1) processes (equation (2)) for the residuals of BTC, ETH, and the CMC200 Index excess returns. Standard errors are reported in parentheses.

BTC. Conversely, when $g_{1,t}$ and $g_{2,t}$ take extreme values but share the same sign, or when they are both close to zero, the investor focuses on the CMC200 Index. In these domains, the investment strategies are conservative and the investor bets on the market by specializing in learning about the market component.

In summary, similar to the case of the stock market, the investor's attention allocation in the cryptocurrency market also involves diversified learning and specialized learning. And the attention strategy matches up with the investment strategy and tends to reduce the uncertainty of the predictors significant in the portfolio. Attention allocation in the two markets differs in terms of learning objectives and learning levels. These differences are driven by the underlying dynamics and learning efficiencies that are unique to each market. For instance, in the cryptocurrency market, the investor may focus more on the systematic predictor due to its higher volatility, whereas in the equity market, the focus is more on the firm-specific predictors. Nevertheless, our investigations imply that the interactions between attention allocation and investment strategies through a feedback mechanism remain consistent across multiple asset types. This emphasizes the robustness of our model in explaining how the investor adapts the strategies based on the market environment.

5.2. *Dynamic attention allocation*

In this subsection, we investigate the dynamic attention and investment behaviors under multiple information sources and multiple assets. Consider a finite investment horizon $t = 0, 1, 2, \ldots, T$. Due to space limitations, we simply assume that the investor's objective is to choose the attention allocation strategy $\boldsymbol{\lambda}$ and the investment strategy $\boldsymbol{\pi}$ that maximize the expected constant absolute risk aversion (CARA) utility at time T. By doing this, we can get rid of the time-inconsistent issue in dynamic mean-variance optimization (Basak and Chabakauri 2010, Wang *et al.* 2021), though our numerical method can also be applied to other popular preference functions and risk metrics (e.g. Zhou *et al.* 2017). We refer the readers to recent work by Cui *et al.* (2024) for the study of pre-committed mean-variance attention and investment strategy. Indeed, in a static model, the CARA utility optimization is equivalent to the mean-variance optimization for normal random payoffs. Specifically, the optimization problem is

$$\max_{(\boldsymbol{\lambda},\boldsymbol{\pi})\in\mathcal{A}_{0,T}} \mathbb{E}\left[-e^{-\gamma W_T}\right], \tag{22}$$

where $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_t\}_{t=0,1,\ldots,T-1}$ and $\boldsymbol{\pi} = \{\boldsymbol{\pi}_{t+}\}_{t=0,1,\ldots,T-1}$. The set of admissible strategies is

$$\mathcal{A}_{0,T} := \left\{ (\boldsymbol{\lambda},\boldsymbol{\pi}) \mid \boldsymbol{\lambda}_t \in \mathbb{R}^{n+1}, \sum_{j\in\mathcal{J}} \lambda_{j,t} = 1, \boldsymbol{\pi}_{t+} \in \mathbb{R}^n, \right. \\ \left. t = 0, 1, \ldots, T-1 \right\}.$$

We denote by $\mathbf{g}_t := \left(g_{1,t}, \ldots, g_{n,t}, g_{m,t}\right)^\top$ and $\mathbf{s}_{t+} := \left(s_{1,t+}, \ldots, s_{n,t+}, s_{m,t+}\right)^\top$ the vector of all the predictive return components and the vector of all the signals, respectively. The Bellman equation for the dynamic problem is

$$V_t(W_t, \mathbf{g}_t) = \max_{\boldsymbol{\lambda}_t} \mathbb{E}_t\left[\max_{\boldsymbol{\pi}_{t+}} \mathbb{E}_{t+}\left[V_{t+1}(W_{t+1}, \mathbf{g}_{t+1})\right]\right], \tag{23}$$

with $V_T(W_T, \mathbf{g}_T) = V_T(W_T) = -\exp(-\gamma W_T)$. We assume the value function is given by the special form: $V_t(w, g) = e^{-\gamma(1+r_f)^{T-t} w} H_t(g)$. Substituting it into (23) and using the dynamic of wealth level (11), we derive

$$H_t(\mathbf{g}_t) = \max_{\lambda_t} \mathbb{E}_t\left[\max_{\pi_{t+}} \mathbb{E}_{t+}\left[e^{-\gamma(1+r_f)^{T-t-1}\boldsymbol{\pi}_{t+}^\top \mathbf{r}_{t+1}} H_{t+1}(\mathbf{g}_{t+1})\right]\right], \tag{24}$$

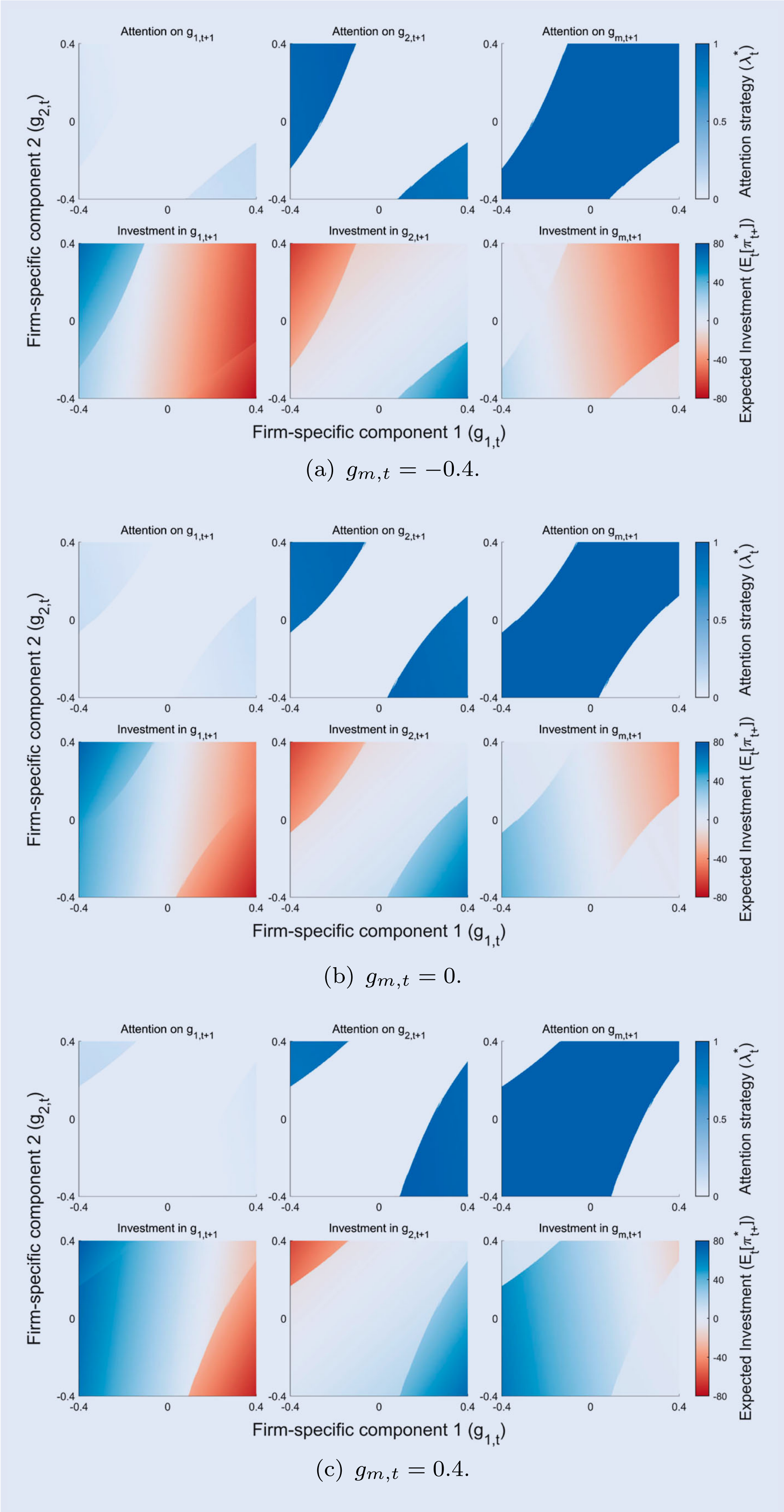


Figure 13. The optimal attention strategies and the expected investment shares for the single-period investment in the cryptocurrency market. (a) $g_{m,t} = -0.4$. (b) $g_{m,t} = 0$ and (c) $g_{m,t} = 0.4$. We choose the systematic component $g_{m,t}$ to be -0.4 (a), its long-term mean 0 (b) and 0.4 (c), respectively. The color gradient in the first row corresponds to changes in the proportion of attention allocation, and the color transition from white to the darkest blue corresponds to a range of proportions from 0 to 1. The color gradient in the second row reflects changes in expected investment in three predictors. The color transition from the darkest red to the darkest blue corresponds to the expected investment in predictors, ranging from deep short to deep long positions.

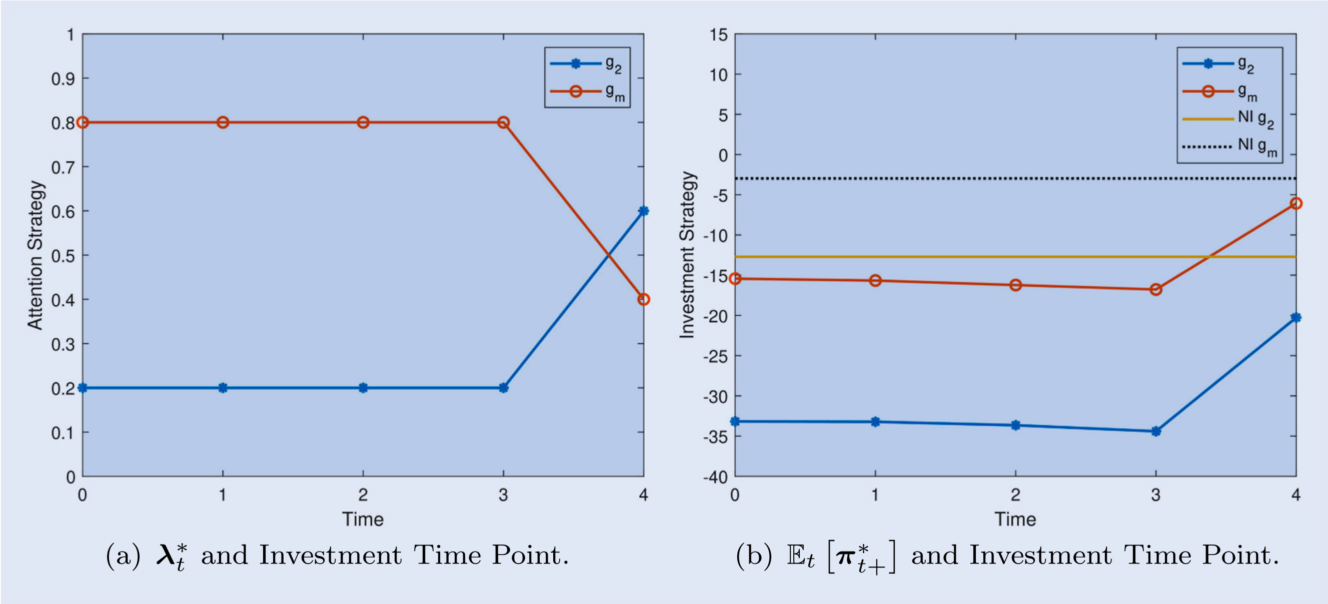


Figure 14. Dynamic evolution of attention strategy and investment strategy over time. (a) $\boldsymbol{\lambda}_t^*$ and Investment Time Point and (b) $\mathbb{E}_t\left[\boldsymbol{\pi}_{t+}^*\right]$ and Investment Time Point. We plot the dynamic evolution of the investor's optimal attention strategy $\boldsymbol{\lambda}_t^*$ and the expected share holdings $\mathbb{E}_t[\boldsymbol{\pi}_{t+}^*]$ in the 5-period investment with 21 attention strategies. The solid yellow and dashed black lines in (b) represent the expected shareholdings when the investor does not learn about the return predictors during the information acquisition phase. The parameters are the same as those used in the single period. The nodes employed for the construction of sparse grids and computation of Gauss-Hermite quadratures are 4, 4, and 5, respectively. The return predictors in the initial investment period are 0, 0.2, and 0, respectively. The monthly risk-free rate $r_f = 0.25\%$.

with $H_T(\mathbf{g}_T) \equiv -1$. To solve the above recursive equations, we propose an efficient numerical method inspired by reinforcement learning techniques. These equations are difficult to solve due to the two-layer optimization in each period and the high dimensionality of the state and strategy vector. Note that the optimal value of the inner optimization problem for $\boldsymbol{\pi}_{t+}$ should be a function of both the return predictors $\mathbf{g}_t \in \mathbb{R}^{n+1}$ and the signal vector $\mathbf{s}_{t+} \in \mathbb{R}^{n+1}$. Therefore, even for the simplest market with two assets, we have a $3 + 2 = 5$-dimensional strategy vector and a $3 + 3 = 6$-dimensional state vector. Indeed, our problem suffers from seriously the so-called *the curse of dimensionality*. To overcome this issue, we propose a numerical method inspired by Q-learning in reinforcement learning (Bertsekas 2019), and approximate the value functions using local sparse grid and Gauss-Hermite quadrature (Brumm *et al.* 2021). The details of the numerical algorithm are in Appendix 2.

Figure 14 shows the dynamic evolution of the investor's optimal attention strategy and the expected investment in three predictors over time. We investigate a 5-period problem in which the investor chooses the attention allocation strategy with each component from a discretized set $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ at each period. In the meanwhile, the investor is still limited by the attention capacity constraint. When constructing sparse grids for the state variables $\mathbf{g}$ and $(\mathbf{g}, \mathbf{s})$ within the algorithm, we choose the numbers of sparse grid nodes to be 4 while computing Gauss-Hermite quadrature for $\mathbf{g}$ or $\mathbf{s}$, we use 5 nodes. The parameter choices are consistent with those adopted in the single-period model. Moreover, we assume that the values of the return predictors are $g_{1,t} = 0$, $g_{2,t} = 0.2$, and $g_{m,t} = 0$ for $t = 0, 1, 2, 3, 4$. We also set the monthly risk-free rate r_f remains to 0.25%.

As shown in figure 14, the investor directs her limited attention primarily between the systematic predictor and the idiosyncratic component of asset 2 (VPU) throughout the entire investment horizon. Specifically, in the initial four investment periods, the investor allocates a significant portion of her attention, 80% of her attention capacity, to monitoring the systematic predictor, despite her awareness that the prior information $g_{2,t}$ has deviated significantly from its long-term mean. It is not until the final period of the investment that the investor changes her consistent decision-making pattern and adopts a 60%: 40% attention allocation to learn about g_2 and g_m. Like the single-period case, the investor never learns about g_1 in the dynamic case. We attribute this inattention to the low learning efficiency of learning about g_1.

Figure 14(b) compares the expected investment of our optimal strategy with the investment strategy when there is no information acquisition. We only report the investment result for two predictors g_2 and g_m since the investor does not learn about g_1 and the investment strategies in g_1 in the two cases are the same. It is clear that, with information acquisition, the investor constructs more aggressive portfolios due to their confidence in predicting asset returns accurately. Moreover, there is a time horizon effect peculiar to the dynamic strategy. That is, the investor tends to trade more aggressively in the first four periods while becoming much more conservative in the last period. This pattern is partially consistent with those in the life-cycle portfolio allocation problem. In standard life-cycle models, younger investors will allocate a larger portion of their investments to relatively riskier stocks, while this allocation becomes progressively more conservative as investors age increases, a phenomenon often referred to as 'age effects' (Viceira 2001, Cocco *et al.* 2005). In our model including information acquisition, the investor shifts from an aggressive strategy to a more cautious strategy in the last period. It's worth noting that, in the static problem, the investor tends to learn about the predictor that has a significant proportion in the portfolio. However, in the dynamic setting, this pattern doesn't hold. By comparing figure 14(a,b), we observe that in the first period, although the investor shorts more of g_2, the attention is more concentrated on the systematic predictor g_m.

It is worth mentioning that in Van Nieuwerburgh and Veldkamp (2010) the investors are expected to be indifferent to attention allocation under the CARA utility. The reasons for the difference between our result and the result in Van Nieuwerburgh and Veldkamp (2010) include that the number of tradable assets is smaller than the number of learning targets in our setup whereas Van Nieuwerburgh and Veldkamp (2010) assumes these two numbers are equal. Moreover, we consider a dynamic problem with mean-reverting return predictors while Van Nieuwerburgh and Veldkamp (2010) investigates a static model. As a result, the perfect substitutions of precision improvements between two assets in Van Nieuwerburgh and Veldkamp (2010) are not valid in our model.

We end this subsection by discussing the scalability of our model for practical use. In the dynamic model, the state variables in our problem include return predictors and signals, which can easily lead to a high-dimensional problem. The attention allocation is made at t while the investment happens at $t+$, indicating the number of the time points of decision-making is twice the horizon T. Therefore, it can be challenging to solve the dynamic problem numerically. However, in practice, the preference of investors like asset managers can be properly characterized by the best mean-variance tradeoff of portfolio return within each period (see Gârleanu and Pedersen 2013, 2016). In this case, the two-phase model in Section 2 is sufficient for practical use and it can be efficiently solved.

When the investor has a relative long-term planning, the Q-learning algorithm applies. Table 3, which is reproduced from Brumm and Scheidegger (2017), compares the number of grid points required for a traditional full grid algorithm and the sparse grid used in our algorithm under different model dimensions. The full grid scales exponentially with increasing dimensions, making it computationally expensive for high-dimension problems. In particular, it requires approximately 0.53 million and 43 million points for the 6- and 8-dimensional cases, respectively. In contrast, the sparse grid method significantly alleviates the curse of dimensionality, reducing the number of points to 137 in the 6-dimensional case and 389 in the 8-dimensional case, thus dramatically improving computational efficiency. We also report the computational time of four models with different dimensions and time points. Note that the number of time points in our problem is twice the horizon length. We run the computations on a personal computer with a 3.60 GHz Intel Core i7-12700KF CPU with 12 cores. Indeed, the speed can be further improved through parallel computing using more cores. For a problem with 6 time points and 8 state variables, our algorithm requires 14.76 hours. The time cost is acceptable when considering the usual practice of portfolio rebalancing such as monthly or quarterly (table 4).

Table 3. Number of points in full grids and sparse grids for different dimensions.

State dimension	$\lvert\mathcal{V}_N^F\rvert$	$\lvert\mathcal{V}_N^S\rvert$
1	9	9
2	81	29
4	6561	137
6	531,441	389
8	43,046,721	849
10	3.4868×10^9	1581

Notes: This table reports the number of points in full grid space ($\lvert\mathcal{V}_N^F\rvert$) and sparse grid space ($\lvert\mathcal{V}_N^S\rvert$) for different dimensions with a fixed maximum refinement level $N = 4$. More details about grid space are given in Appendix 2.

Table 4. Computation time for different state dimensions and number of time points.

	Number of time points	
State dimension	6	8
6	2.47	3.09
8	14.76	33.29

Notes: This table reports the computation time (h) for different state dimensions and the number of time points.

5.3. *Discussions on asset pricing*

In this subsection, we use the CAPM theory to investigate the economic equilibrium within our model. Following Andrei *et al.* (2023), we are particularly interested in asset price changes before and after a representative investor observes the signals. We let the risk-free rate $r_f = 0$ and assume there are only three dates $t \in \{0, 0+, 1\}$. Recall that the investor's decisions are made according to figure 1 with $t = 0$. Let $\mathbf{p}_{0+} = (p_{1,0+}, p_{2,0+}, \ldots, p_{n,0+})^\top$ be the equilibrium price vector of the n risky assets when the vector of signals $\mathbf{s}_{0+}$ is accessible.

Suppose there are noise traders respectively at time 0, 0+ who have inelastic demands of $\mathbf{x}_0$, $\mathbf{x}_{0+}$ shares with zero means. The components of $\mathbf{x}_0$, $\mathbf{x}_{0+}$ are assumed to be independent of each other as well as the other random variables in our model. The market portfolio is denoted by $\mathbf{M} \in \mathbb{R}^n$, whose components M_is are positive and satisfies $\sum_{i=1}^n M_i = 1$. We assume a representative mean-variance investor who allocates attention to $n+1$ signal sources using the strategy $\boldsymbol{\lambda}_0^*$. Therefore, the total supply of assets available to the mean-variance investor is $\mathbf{M} - \mathbf{x}_0$ at time 0 and $\mathbf{M} - \mathbf{x}_0 - \mathbf{x}_{0+}$ at time 0+.

Define the dollar excess return between 0 and 0+ as $\mathbf{r}_{0+}^g := \mathbf{p}_{0+} - \mathbf{p}_0$ and the market excess return as $r_{0+}^m := \mathbf{M}^\top \mathbf{r}_{0+}^g$. The following proposition illustrates a CAPM relation and the market risk premium during the announcement of signals. Note that for simplicity, we suppress the time subscript '0' in $\boldsymbol{\Sigma}_0(\cdot)$ and $\boldsymbol{\lambda}_0^*$. The proofs of this proposition are in Appendix A.7.

PROPOSITION 5.1 (CAPM) *The following relation holds when the signals about return predictors are accessible:*

$$\mathbb{E}\left[\mathbf{r}_{0+}^g\right] = \boldsymbol{\beta}\mathbb{E}\left[r_{0+}^m\right], \tag{25}$$

with

$$\boldsymbol{\beta} = \frac{(\boldsymbol{\Sigma}(\mathbf{0}) - \boldsymbol{\Sigma}(\boldsymbol{\lambda}^*))\,\mathbf{M}}{\mathbf{M}^\top(\boldsymbol{\Sigma}(\mathbf{0}) - \boldsymbol{\Sigma}(\boldsymbol{\lambda}^*))\,\mathbf{M}}, \tag{26}$$

where the market risk premium is given by

$$\mathbb{E}\left[\mathbf{r}^m_{0+}\right] = \gamma\mathbf{M}^\top\boldsymbol{\Sigma}(\mathbf{0})\mathbf{M} - \gamma\mathbf{M}^\top\boldsymbol{\Sigma}(\boldsymbol{\lambda}^*)\mathbf{M}. \tag{27}$$

Proposition 5.1 describes a CAPM relation that highly resembles that in Andrei *et al.* (2023). However, there are notable differences between their findings and ours. In Andrei *et al.* (2023), there is no information capacity constraint on their attention strategy, and their focus is primarily on firm-specific announcements. In contrast, our attention strategy is subject to a capacity constraint. When the investor becomes more attentive toward one signal resource, the attention paid to the other resources must decrease and thus the risk premium and beta of any related assets can be affected. This differs from Andrei *et al.* (2023) where increasing attention on one asset only affects that asset itself. Moreover, our setting incorporates the proportion of attention allocated to the systematic news, i.e. λ^*_m. Changes in λ^*_m can affect all the risky assets traded in the market.

The quantity $\gamma\mathbf{M}^\top\boldsymbol{\Sigma}(\mathbf{0})\mathbf{M}$ measures a market-wide prior uncertainty of the assets payoffs. Without attention ($K = 0$) or when it is extremely difficult to extract information from the signals ($\boldsymbol{\theta} = \mathbf{0}$), the market risk premium becomes zero. This implies that holding the market portfolio from time t to time $t+$ entails no risk. In other words, the market risk premium increases and rewards the investors for resolving uncertainty through allocating attention. A higher market risk premium also implies a steeper security market line (SML) on announcement days. This theoretical prediction aligns with the findings in Andrei *et al.* (2023) and has been empirically validated in their real-market analysis.

Proposition 5.1 also has predictions on the firms' beta. Since $\mathbf{M}$ represents the market portfolio, we impose a reasonable assumption $\mathbf{b}^\top\mathbf{M} = 1$. Then we obtain

$$\beta_i = \frac{M_i\omega_i^2(1 - e^{-\theta_i\lambda_i^*K}) + b_i\omega_m^2(1 - e^{-\theta_m\lambda_m^*K})}{\sum_{l=1}^n M_l^2\omega_l^2(1 - e^{-\theta_l\lambda_l^*K}) + \omega_m^2(1 - e^{-\theta_m\lambda_m^*K})}. \tag{28}$$

Suppose $b_i > 0$. Mathematically, we find that β_i increases as we reduce λ_l^* for $l \neq i$ and improve λ_i^*. This relationship implies that the firm i's beta becomes stronger when the investor shifts attention from other firms to firm i. Moreover, as the investor focuses more on the systematic return predictor ($\lambda_m^* \to 1$), less attention is paid to the remaining predictors and then the firm's market beta β_i converges to b_i.

As $\theta_i = 0$ can be interpreted as the firm i having no announcement at time $0+$, our model also implies an improved risk premium on announcement days. This provides an explanation for the documented earning announcement premium (see, e.g. Ball and Kothari 1991, Cohen *et al.* 2007, Andrei *et al.* 2023).

6. Conclusion

In this paper, we investigate the investor's attention allocation and portfolio selection in a multi-asset incomplete market with return predictability. Utilizing the entropy-based information capacity and Bayes learning, we model the information acquisition procedure of a mean-variance investor confronted with multiple information sources at both firm-specific and systematic levels. We show theoretically that the investor's optimal attention allocation problem is to maximize the expected squared posterior Sharpe ratio and this objective can be decomposed into two parts including the alpha payoffs from resolving the uncertainty of firm-specific components and the beta payoffs from resolving uncertainty of the systematic component.

Our results show that the investor tends to allocate attention to the return predictors taking extreme values, leading to a U-shaped relationship between the attention allocation and the predictor level. Moreover, there exists competition among the return predictors to attract the investor's attention, as a significant deviation from the predictor's long-term mean will motivate the investor to focus on it and become inattentive to the others. Our incomplete market framework enables us to derive analytical results in specific cases. These results illustrate that the investor will engage in either diversified learning or specialized learning, depending on factors such as the uncertainty of components, the risk exposure of asset returns to the systematic component, learning efficiency, and information capacity. We investigate the investor's benefit from learning defined as the improvements in the objective function resulting from learning. The results demonstrate the crucial role of attention allocation in effective risk management in response to volatile market conditions while optimizing portfolios.

The extensions to the emerging cryptocurrency market further confirm the patterns of switching between specialized and diversified learning and adjusting investment strategies for alpha or beta payoffs accordingly, while revealing differences in learning objectives and levels depending on specific market conditions. We also study the multi-period problems using Q-learning in the numerical study. The results demonstrate that the optimal attention allocation strategy always matches up with the investment strategy, emphasizing a significant feedback effect emanating from the attention strategy to investment. Our paper further sheds light on the time-horizon effects in multi-period learning and investment patterns that the investor exhibits a notably stronger tendency towards learning about and trading on a particular return predictor in preceding periods than in the terminal period. Finally, we also study the equilibrium price in the context of attention allocation and derive the CAMP relation in a two-period economy. We find that the investor's attention allocation has a significant impact on the firms' betas, and thus on the market risk premium and the SML for the firm on the signal-announcement day.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China (72301106, 71971083, 71931004), Shanghai Pujiang Program (22PJC038), Fundamental Rese arch Funds for the Central Universities.

Supplemental data

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/14697688.2024.2423702.

ORCID

Yun Shi http://orcid.org/0000-0002-4746-2312
Tingjin Yan http://orcid.org/0000-0002-3621-9042
Yong Zhou http://orcid.org/0000-0001-9751-3141

References

Abel, A.B., Eberly, J.C. and Panageas, S., Optimal inattention to the stock market with information costs and transactions costs. *Econometrica*, 2013, **81**, 1455–1481.
Andrei, D. and Hasler, M., Dynamic attention behavior under return predictability. *Manage. Sci.*, 2020, **66**, 2906–2928.
Andrei, D., Friedman, H. and Ozel, N.B., Economic uncertainty and investor attention. *J. Financ. Econ.*, 2023, **149**, 179–217.
Ball, R. and Kothari, S.P., Security returns around earnings announcements. *Account. Rev.*, 1991, **66**, 718–738.
Barber, B.M. and Odean, T., All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Rev. Financ. Stud.*, 2008, **21**, 785–818.
Barber, B.M., Lin, S. and Odean, T., Resolving a paradox: Retail trades positively predict returns but are not profitable. *J. Financ. Quant. Anal.*, 2024, **59**, 2547–2581. DOI: 10.1017/S0022109023000601.
Basak, S. and Chabakauri, G., Dynamic mean-variance asset allocation. *Rev. Financ. Stud.*, 2010, **23**, 2970–3016.
Ben-Rephael, A., Da, Z. and Israelsen, R.D., It depends on where you search: Institutional investor attention and underreaction to news. *Rev. Financ. Stud.*, 2017, **30**, 3009–3047.
Bertsekas, D., *Reinforcement Learning and Optimal Control*, 2019 (Athena Scientific: Nashua, NH).
Brumm, J. and Scheidegger, S., Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica*, 2017, **85**, 1575–1612.
Brumm, J., Krause, C., Schaab, A. and Scheidegger, S., Sparse grids for dynamic economic models. Available at SSRN 3979412, 2021.
Campbell, J.Y., *Financial Decisions and Markets: A Course in Asset Pricing*, 2017 (Princeton University Press: Princeton, NJ).
Chabris, C. and Simons, D., *The Invisible Gorilla: How Our Intuitions Deceive Us*, 2011 (Random House Inc.: New York).
Chen, J., Tang, G., Yao, J. and Zhou, G., Investor attention and stock returns. *J. Financ. Quant. Anal.*, 2022, **57**, 455–484.
Cocco, J.F., Gomes, F.J. and Maenhout, P.J., Consumption and portfolio choice over the life cycle. *Rev. Financ. Stud.*, 2005, **18**, 491–533.
Cohen, D.A., Dey, A., Lys, T.Z. and Sunder, S.V., Earnings announcement premia and the limits to arbitrage. *J. Account. Econ.*, 2007, **43**, 153–180.
Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, 1991 (Wiley-Interscience: New York).
Cui, X., Gao, J. and Kong, L., Limited attention allocation in a stochastic linear quadratic system with multiplicative noise. arXiv:2403.18528, 2024.
Da, Z., Engelberg, J. and Gao, P., In search of attention. *J. Finance*, 2011, **66**, 1461–1499.
Daniel, K., Hirshleifer, D. and Subrahmanyam, A., Investor psychology and security market under-and overreactions. *J. Finance*, 1998, **53**, 1839–1885.
Gabaix, X., Chapter 4 - Behavioral inattention. In *Handbook of Behavioral Economics - Foundations and Applications 2*, Vol. 2, pp. 261–343, 2019 (North-Holland: Amsterdam).
Gârleanu, N. and Pedersen, L.H., Dynamic trading with predictable returns and transaction costs. *J. Finance*, 2013, **68**, 2309–2340.
Gârleanu, N. and Pedersen, L.H., Dynamic portfolio choice with frictions. *J. Econ. Theory.*, 2016, **165**, 487–516.
Gervais, S. and Odean, T., Learning to be overconfident. *Rev. Financ. Stud.*, 2001, **14**, 1–27.
Glasserman, P. and Mamaysky, H., Investor information choice with macro and micro information. *Rev. Asset Pric. Stud.*, 2023, **13**, 1–52.
Gondhi, N., Rational inattention, misallocation, and the aggregate economy. *J. Monet. Econ.*, 2023, **136**, 50–75.
Hou, K., Peng, L. and Xiong, W., A tale of two anomalies: The implications of investor attention for price and earnings momentum. Available at SSRN 976394, 2009.
Jacobs, H. and Weber, M., Losing sight of the trees for the forest? Attention allocation and anomalies. *Quant. Finance*, 2016, **16**, 1679–1693.
Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L., Time-varying fund manager skill. *J. Finance*, 2014, **69**, 1455–1484.
Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L., A rational theory of mutual funds' attention allocation. *Econometrica*, 2016, **84**, 571–626.
Kahneman, D., *Attention and Effort*, 1973 (Englewood Cliffs, NJ: Prentice-Hall).
Liu, Y. and Tsyvinski, A., Risks and returns of cryptocurrency. *Rev. Financ. Stud.*, 2021, **34**, 2689–2727.
Liu, Y., Tsyvinski, A. and Wu, X., Common risk factors in cryptocurrency. *J. Finance*, 2022, **77**, 1133–1177.
Liu, H., Peng, L. and Tang, Y., Retail attention, institutional attention. *J. Financ. Quant. Anal.*, 2023, **58**, 1005–1038.
Luo, Y., Robustly strategic consumption–portfolio rules with informational frictions. *Manage. Sci.*, 2017, **63**, 4158–4174.
Ma, G. and Zhu, S.P., Optimal investment and consumption under a continuous-time cointegration model with exponential utility. *Quant. Finance*, 2019, **19**, 1135–1149.
Mondria, J., Portfolio choice, attention allocation, and price comovement. *J. Econ. Theory*, 2010, **145**, 1837–1864.
Peng, L., Learning with information capacity constraints. *J. Financ. Quant. Anal.*, 2005, **40**, 307–329.
Peng, L. and Xiong, W., Investor attention, overconfidence and category learning. *J. Financ. Econ.*, 2006, **80**, 563–602.
Poterba, J.M. and Summers, L.H., Mean reversion in stock prices: Evidence and implications. *J. Financ. Econ.*, 1988, **22**, 27–59.
Reis, R., Inattentive consumers. *J. Monet. Econ.*, 2006, **53**, 1761–1800.
Shannon, C.E., A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423.
Sims, C.A., Implications of rational inattention. *J. Monet. Econ.*, 2003, **50**, 665–690.
Van Nieuwerburgh, S. and Veldkamp, L., Information acquisition and under-diversification. *Rev. Econ. Stud.*, 2010, **77**, 779–805.
Veldkamp, L., Information markets and the comovement of asset prices. *Rev. Econ. Stud.*, 2006, **73**, 823–845.
Viceira, L.M., Optimal portfolio choice for long-horizon investors with nontradable labor income. *J. Finance*, 2001, **56**, 433–470.
Wang, T., Jin, Z. and Wei, J., Mean-variance portfolio selection with non-negative state-dependent risk aversion. *Quant. Finance*, 2021, **21**, 657–671.
Yan, T., Chiu, M.C. and Wong, H.Y., Pairs trading under delayed cointegration. *Quant. Finance*, 2022, **22**, 1627–1648.
Zhang, Y., Jin, Z., Wei, J. and Yin, G., Mean–variance portfolio selection with dynamic attention behavior in a hidden Markov model. *Automatica*, 2022, **146**, 110629.
Zhou, K., Gao, J., Li, D. and Cui, X., Dynamic mean–VaR portfolio selection in continuous time. *Quant. Finance*, 2017, **17**, 1631–1643.