INVESTOR ATTENTION, OVERCONFIDENCE AND CATEGORY LEARNING

Lin Peng
Wei Xiong

Investor Attention, Overconfidence and Category Learning
Lin Peng and Wei Xiong
NBER Working Paper No. 11400
June 2005
JEL No. G0, G1

ABSTRACT

Motivated by psychological evidence that attention is a scarce cognitive resource, we model investors' attention allocation in learning and study the effects of this on asset-price dynamics. We show that limited investor attention leads to ``category-learning'' behavior, i.e., investors tend to process more market and sector-wide information than firm-specific information. This endogenous structure of information, when combined with investor overconfidence, generates important features observed in return comovement that are otherwise difficult to explain with standard rational expectations models. Our model also demonstrates new cross-sectional implications for return predictability.

Lin Peng
Baruch College
City University of New York
New York, NY 10010
lin_peng@baruch.cuny.edu

Wei Xiong
Princeton University
Bendheim Center for Finance
26 Prospect Avenue
Princeton, NJ 08450
and NBER
wxiong@princeton.edu

1 Introduction

Standard asset-pricing models are typically based on the assumption that markets distill new information with lightning speed and that they provide the best possible estimate of all asset values. In reality, this distillation and estimation requires investors' close attention to processing information and to incorporating this knowledge into their decisions. Several recent studies suggest that investor attention may play an important role in determining asset prices. Important news or information is not reflected by prices until investors pay attention to it.[1] Despite the growing empirical evidence, there has been little formal analysis on this issue. In this paper, we study the effects of investor attention on asset price dynamics.

We emphasize that attention is a scarce cognitive resource (Kahneman, 1973). Attention to one task necessarily requires a substitution of cognitive resources from other tasks. When it comes to investment decisions, given the vast amount of information available and the inevitability of limited attention, investors have to be selective in information processing.

We consider a discrete-time model with an infinite number of periods. There are multiple risky assets and one risk-free asset in the economy. The risk-free asset offers a constant rate of return. The dividend payoff of each risky asset is determined by a linear combination of three unobservable random factors with independent Gaussian distributions: market, sector, and firm-specific.

We study the learning process of a representative investor who has limited attention. Facing unobservable fundamental factors in her portfolio, the investor processes information to infer their values. We view this representative investor as one of many retail investors in the stock market who face similar uncertainty in their portfolios. They access similar information sources, such as newspapers, analyst reports and media coverage. These investors are also subject to similar attention constraints and behavioral biases in information processing.

We model the investor's learning process as follows. The investor first generates a vector of

---

[1]Huberman and Regev (2001) provide a vivid example: the publication of an article in the *New York Times* about a new cancer curing drug from EntreMed attracted great public attention and generated a daily return of more than 300% in its stocks, even though the same story had already been published more than five months earlier in *Nature* and other newspapers. Other studies, e.g., Hirshleifer, Hou, Teoh, and Zhang (2004), Hong, Walter, and Valkanov (2003), Hou and Moskowitz (2003), and DellaVigna and Pollet (2003), provide evidence that stock prices do not fully incorporate all information that appears in the public domain, such as the prices of other assets and certain variables in firms' financial statements.

signals through information processing. This process is affected by the investor's attention constraint and her attention allocation. We adopt the entropy concept from information theory to measure information and impose the attention constraint as the maximum amount of information that the investor can process each period. Given the multiple sources of uncertainty, the investor optimally allocates her attention across them. As the investor allocates more attention to a factor, she processes more information. After gathering the signals, the investor then incorporates them into her beliefs through Bayesian updating.

To maximize her expected lifetime utility, the investor optimally makes her consumption decisions based on her beliefs about fundamental factors. An exponential utility function and Gaussian distributions for all variables allow a linear equilibrium, in which asset prices are determined through the pricing kernel determined by the investor's marginal utility of consumption. In the equilibrium, the investor allocates attention across fundamental factors to reduce the total uncertainty of her portfolio while asset prices fluctuate as the investor updates her beliefs based on the processed information. In this way, the investor's attention allocation affects the asset-price dynamics.

Our model shows that limited attention leads to "category-learning" behavior: an attention-constrained investor tends to allocate more attention to market- and sector-level factors than to firm-specific factors. In severely constrained cases, the investor allocates all attention to market- and sector-level information and ignores all the firm-specific data. For instance, during the internet bubble period, firms that had changed to dot.com names without any fundamental changes in strategies earned significant abnormal returns around their name-change announcements (Cooper, Dimitrov, and Rau, 2001). This example shows how inattentive investors could be to firm-specific information.

The endogenous information structure derived from investors' attention allocation is particularly useful in studying the interaction between investors' attention and their biased reactions to information. Several recent studies suggest that biased reactions to information provide helpful insights in understanding many empirical anomalies that have been discovered over the past two decades.[2] Since biased reactions only occur when investors attend to

---

[2]See Hirshleifer (2001) and Barberis and Thaler (2003) for recent reviews of this literature. For example, in explaining overreaction and underreaction of stock prices in different situations, Daniel, Hirshleifer, and Subrahmanyam (1998) analyze overconfidence and self-attribution bias in investors' responses to information, while Barberis, Shleifer,

certain pieces of information and/or ignore others, the attention allocation decisions studied in our model determine the cross-sectional patterns of these biased reactions.

We give special consideration to one specific form of investor bias, overconfidence. Experimental studies have shown that the trait of overconfidence is particularly severe in those faced with diffuse tasks that require difficult judgements but provide only noisy and delayed feedback (see Einhorn, 1980). The fundamental valuation of financial securities is a good example of this type of difficult tasks, one which becomes even more challenging when investors have limited attention. We model overconfidence as the investor's exaggeration of her information processing ability. As a result, the investor overestimates the precision of her information, in a way consistent with other overconfidence models in the literature.[3]

Our model captures three features of asset return comovement observed by recent empirical studies. First, return correlations between firms can be higher than their fundamental correlations. This result is generated by the interaction of the investor's category learning behavior (her tendency to processing more market- and sector-level information) to her overreaction to the processed information. This result is supported by the empirical studies of Shiller (1989) and Pindyck and Rotemberg (1993) on the comovement between U.K. and U.S. stock markets and the comovement of individual U.S. stocks.

Second, our model shows that across different sectors, there is a negative relation between the average return correlation of firms in a sector and their stock price informativeness. For a sector with a higher information-processing efficiency, rather than treating the sector as a category, the investor allocates relatively more attention to firm-specific information. Consequently, these firms' stock prices will be more informative about their future fundamentals, and their returns will have relatively more firm-specific variation (or smaller correlations). This result provides an explanation to the findings by Morck, Yeung, and Yu (2000) and Durnev, Morck, Yeung, and Zarowin (2003) that stock returns are more informative about changes in future earnings in industries or countries with less correlated stock returns.

Third, our model shows that as information technology advances over time, investors'

---

and Vishny (1998) consider investor representativeness and conservatism. Along another line, Hong and Stein (1999) and Hirshleifer and Teoh (2004) analyze models in which some useful public information is either ignored or only gradually recognized by investors.

[3]See, for example, Kyle and Wang (1997), Daniel, Hirshleifer, and Subrahmanyam (1998), Odean (1998), Bernardo and Welch (2001), Gervais and Odean (2001), and Scheinkman and Xiong (2003).

attention constraints become less binding and they can allocate relatively more attention to firm-specific information, thereby reducing return correlations. This result explains the finding of Campbell, Lettau, Malkiel, and Xu (2001), who document a decreasing trend in the return correlation of U.S. stocks over the last thirty years.

It may appear that our model's implications for return comovement can also be derived from a rational expectations model of costly information acquisition. However, West (1988) analyzes a rational expectations model and shows that improved information about future cashflows actually decreases return volatility. The basic intuition is that more information only allows a rational investor to resolve uncertainty earlier, but does not increase the level of return variation. Thus, it is difficult for standard rational expectations models to explain the empirical evidence on return comovement based on cross-sectional difference and time-trend in information cost.

The investor's attention allocation decisions also directly affect the cross-sectional patterns of asset-return predictability. When the investor allocates more attention to individual firms in a sector and processes more firm-specific information, there will be more pronounced overreaction-driven predictability in firm-specific returns. In the meantime, more firm-specific information processed leaves an ignored public signal less valuable in predicting firms' future returns. Since the investor's attention allocation to individual firms in a sector is negatively related to the average return correlation in the sector, our model provides two new testable implications: after controlling for the degree of investor overconfidence, firms in a sector with a lower average return correlation tend to have more pronounced overreaction-driven return predictability (e.g., long-run price reversals and short-term price momentum); on the other hand, an ignored public signal (such as certain variables in firms' financial statements) will have less predictive power for these firms' returns.

The paper is organized as follows. Section 2 reviews the related studies on attention. We introduce the model and derive the equilibrium in Section 3. Section 4 discusses the investor's category learning behavior under attention constraints. In Section 5, we describe the cross-sectional and time series implications of our model for asset return comovement. Section 6 illustrates the implications of this on return predictability. Section 7 concludes. The Appendix provides technical derivations and proofs.

2 Related studies on attention

There is a large body of psychological research on human attention. These studies suggest that people's ability to simultaneously perform different tasks depends on whether they involve only perceptual analysis or more central cognitive analysis requiring memory retrieval and action planning. Although it is possible for people to simultaneously handle multiple perceptual tasks, such as typing while listening to music, psychological evidence shows that overlap in the central cognitive operations of different tasks does not occur successfully, except in a few special cases.

Pashler and Johnston (1998) summarize various supporting evidence that there is a limit to the central cognitive processing capacity of the human brain. The operation of human brains is intuitively described by psychologists as similar to that of a single-processor computers. Both deal with multiple tasks by working on one task at a time, alternating between these tasks in order to respond to inputs in a timely fashion. The rate or efficiency of processing for each task depends on the processing time the computer allocates to the task. This is the basic concept that we adopt to model information processing by investors.

Recent theoretical work in economics and finance has begun to explore some of the consequences of limited investor attention. Hirshleifer and Teoh (2003) analyze firms' choices between alternative methods for presenting information, and the effects of different presentations on market prices when investors have limited attention and processing power. Hirshleifer, Lim, and Teoh (2003) address firms' incentives to withhold information from credulous and inattentive investors. Hirshleifer and Teoh (2004) provide a model of return predictability based on inattentive investors' negligence of current earnings, different earnings components, or information in investment. Our model also addresses the effect of neglected information on asset returns. In particular, we derive new cross-sectional implications linking the strength of this effect to return correlation.

Sims (2003) adopts the concept of channel capacity from information theory to study information processing constraints in a dynamic control problem without financial assets. Peng (2005) studies an information capacity constraint in investors' learning processes. Van Nieuwerburgh and Veldkamp (2004) discuss portfolio under-diversification caused by in-

vestors' learning constraints. We use an information measure shared by these studies, and focus on investors' category learning behavior and its relationship to return comovement and predictability.

Psychological studies, as reviewed by Yantis (1998), suggest that attention can be directed by people's deliberate strategies and intentions. Gabaix and Laibson (2003) analyze a model of directed attention of economic agents who allocate thinking time to choose a consumption good from several alternatives, and Gabaix et al. (2003) provide some further experimental evidence. Our model also focuses on the effects of actively controlled investor attention.[4]

Investors' limited attention or computational capacity also motivates several recent studies of heuristics that simplifies problem-solving, e.g., Barberis and Shleifer (2003), Mullainathan (2002), and Hong and Stein (2003). In particular, Barberis and Shleifer analyze the effects of investors' style strategies, i.e., investors allocate investment based on exogenous asset styles and simultaneously move in and out of a style depending on its recent performance, on excessive comovement among assets of the same style. In our model, investors with limited attention form asset categories based on their fundamentals, and the excessive comovement is driven by investors' inattention to firm-specific information. This framework allows us to draw explicit links between the dynamics of information flow and the dynamics of return comovement and predictability.

3 The model

We develop a model with a infinite number of periods, $t = 1, 2, 3, ..., \infty$. There is a representative investor who invests in a portfolio of risky financial assets. We allow the investor to borrow and lend at a constant risk-free rate of r. We view the representative investor in our model as one of many retail investors in the stock market who face similar uncertainty in their portfolios and who access similar information sources such as newspapers, analyst reports, and media coverage. These investors are also subject to similar behavioral biases and attention constraints in their learning processes.

---

[4]Attention can sometimes be captured by an abrupt onset of stimulus and other salient events. Barber and Odean (2003) examine the stock trading generated by investor attention that is driven by salient events.

3.1   Investor preference and factor structure of asset fundamentals

The investor has an exponential utility function of consumption:

$$u(c) = -\frac{1}{\gamma} e^{-\gamma c} \tag{1}$$

where γ is her absolute risk-aversion coefficient. Each period, the investor chooses a consumption level to maximize her expected lifetime utility:

$$\max \ E_t \left[\sum_{s=t}^{\infty} \delta^{s-t} u(c_s) \right]. \tag{2}$$

$\delta \in (0, 1)$ is the time preference parameter, and c_s is the consumption choice in period s.

The investor holds a portfolio that spans m sectors, with n firms in each sector. We view a sector as an industry or nation. Each firm pays a dividend every period. We denote the dividend from the j-th firm of the i-th sector in period t by $d_{i,j,t}$, and summarize the dividends of all assets by a vector

$$D_{mn\times 1}(t) = (d_{1,1,t}, \cdots, d_{1,n,t}, \cdots, d_{i,j,t}, \cdots, d_{m,1,t}, \cdots, d_{m,n,t})^T \tag{3}$$

where 'T' is the transpose operator.

The dividends are linear combinations of random fundamental factors:

$$d_{i,j,t} \ = \ h_t + f_{i,t} + g_{i,j,t}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, n \tag{4}$$

where h_t is a market factor, $f_{i,t}$ is the common factor for sector i, and $g_{i,j,t}$ is the firm-specific factor for the j-th firm in the i-th sector. These factors are unobservable and independent of each other. Their distributions are known to the investor. We assume that these factors are identically and independently distributed across periods:

$$h_t \sim N(\bar{h}, \sigma_h^2), \tag{5}$$

$$f_{i,t} \sim N(\bar{f}, \sigma_f^2), \quad i = 1, \cdots, m \tag{6}$$

$$g_{i,j,t} \sim N(\bar{g}, \sigma_g^2), \quad i = 1, \cdots, m, \quad j = 1, \cdots, n. \tag{7}$$

In this specification, the market factor has a Gaussian distribution with a mean of $\bar{h}$ and a variance of σ_h^2; all sector factors have the same Gaussian distribution with a mean of $\bar{f}$ and

a variance of σ_f^2; and all firm-specific factors have an identical Gaussian distribution with a mean of $\bar{g}$ and a variance of σ_g^2. While we make these specific assumptions to simplify our analysis, they are not critical to our main results.

3.2   The learning process

Although the fundamental factors are unobservable, the investor is able to analyze them one period before their realization in dividends.[5] A more precise forecast of future dividends benefits the investor by allowing her to make more efficient intertemporal consumption decisions.

We emphasize two important features in the investor's learning process: limited attention and overconfidence. The investor's limited attention determines the maximum amount of information she can process each period. We treat learning about each of the independent fundamental factors as separate tasks, similar to the way that English and Mathematics are studied separately in school. The investor allocates her attention across these tasks.

We use κ to represent the limit in attention. We denote the fraction of attention allocated to the market factor h_t by $\lambda_{h,t} \in [0,1]$, to a sector-wide factor f_i by $\lambda_{f,i,t} \in [0,1]$, and to a firm-specific factor $g_{i,j}$ by $\lambda_{g,i,j,t} \in [0,1]$. We summarize the allocations in a vector:

$$\Lambda_t = (\lambda_{h,t}, \lambda_{f,1,t}, \cdots, \lambda_{f,m,t}, \lambda_{g,1,1,t}, \cdots, \lambda_{g,i,j,t}, \cdots, \lambda_{g,m,n,t})^T. \tag{8}$$

The allocations should sum up to be less than or equal to one:

$$\lambda_{h,t} + \sum_{i=1}^{m} \lambda_{f,i,t} + \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{g,i,j,t} \leq 1. \tag{9}$$

The investor's learning process involves first collecting firms' information in the public domain, such as news releases, media coverage, and analyst reports. Then she forms her own views. To abstract from these complex activities involved in actual learning processes, the investor in our model devotes her attention to processing information, extracts a noisy signal about the underlying factor, and uses the signal to update her belief according to the

---

[5]If we were to allow the investor to analyze future fundamental factors several periods ahead, doing so would add more learning tasks to the investor's attention allocation, since the investor can choose to analyze factors of future periods. However, this extension should not affect the relative allocation across factors in a stationary equilibrium.

Bayes rule. We denote $s_{h,t}$, $s_{f,i,t}$ and $s_{g,i,j,t}$ as the extracted signals for the market, sector, and firm-specific factors, respectively:

$$s_{h,t} = h_{t+1} + \epsilon_{h,t}, \quad \epsilon_{h,t} \sim N(0, \eta_{h,t}^2), \tag{10}$$

$$s_{f,i,t} = f_{i,t+1} + \epsilon_{f,i,t}, \quad \epsilon_{f,i,t} \sim N(0, \eta_{f,i,t}^2), \quad i = 1, \cdots, m \tag{11}$$

$$s_{g,i,j,t} = g_{i,j,t+1} + \epsilon_{g,i,j,t}, \quad \epsilon_{g,i,j,t} \sim N(0, \eta_{g,i,j,t}^2), \quad i = 1, \cdots, m, \ j = 1, \cdots, n \tag{12}$$

where the signal noises $\epsilon_{h,t}$, $\epsilon_{f,i,t}$ and $\epsilon_{g,i,j,t}$ are assumed to have Gaussian distributions with zero means and variances of $\eta_{h,t}^2$, $\eta_{f,i,t}^2$ and $\eta_{g,i,j,t}^2$, respectively. The signal noises are independent of each other, since they are generated through separate tasks. Their variances depend on the investor's attention allocations. We summarize all the signals in a vector S_t.

If the investor devotes more attention to one factor, the extracted signal will become more precise. Following Sims (2003), we use the concept of entropy from information theory to measure the amount of information about the underlying factor contained in a signal. Appendix A provides an introduction to the entropy measure. Basically, the entropy of a random variable measures its uncertainty. In the case of the market factor h_{t+1}, which has a Gaussian distribution $N(0, \sigma_h^2)$, its entropy, H, is determined by its variance σ_h^2:

$$H(h_{t+1}) = \frac{1}{2} \log \sigma_h^2 + 0.5 \log(2\pi e), \tag{13}$$

where π is Archimedes' constant and e is Euler's constant. This expression is intuitive, since for a Gaussian distribution, more uncertainty is associated with a larger variance parameter. Given the signal, $s_{h,t}$, the investor updates her belief about the underlying factor, h_{t+1}, according to the Bayes rule and forms a posterior belief:

$$h_{t+1}|s_{h,t} \sim N(\hat{h}_t, \hat{\omega}_{h,t}^2), \tag{14}$$

where $\hat{h}_t$ is the conditional mean and $\hat{\omega}_{h,t}^2$ is the conditional variance.

Information theory defines the amount of information contained in $s_{h,t}$ about h_{t+1}, $I(h_{t+1}; s_{h,t})$, as the reduction in the entropy (uncertainty) of h_{t+1} due to the knowledge of $s_{h,t}$:

$$I(h_{t+1}; s_{h,t}) \equiv H(h_{t+1}) - H(h_{t+1}|s_{h,t}) = \frac{1}{2} \log \left(\sigma_h^2 / \hat{\omega}_{h,t}^2 \right). \tag{15}$$

Since the signal $s_{h,t}$ is the output of the investor's information processing, its information content is determined by the investor's attention input to generating the signal. We assume a linear relation between these two variables:

$$I(h_{t+1}; s_{h,t}) = \frac{1}{2}\theta_h \kappa \lambda_{h,t}, \tag{16}$$

where $\kappa \lambda_{h,t}$ is the amount of attention allocated to the market factor, and $\theta_h > 0$ is an information-efficiency parameter that measures the efficiency of the investor's information processing of the market factor. For a given value of θ_h, as the investor allocates more attention to the factor, she processes more information about it.

We allow for the possibility that the investor is overconfident about her information-processing ability. We assume that the investor perceives that the amount of information in her signal $s_{h,t}$ is not exactly given by equation (16), but by

$$\hat{I}(h_{t+1}; s_{h,t}) = \frac{1}{2}\phi\theta_h \kappa \lambda_{h,t}, \tag{17}$$

where she exaggerates the amount of information by an overconfidence parameter $\phi \geq 1$. If $\phi > 1$, the investor is overconfident about her information-processing ability. If ϕ is equal to one, we obtain the unbiased case.

Equation (17) implies that the investor may underestimate the variance of her signal noise as $\hat{\eta}_{h,t}^2$, instead of $\eta_{h,t}^2$. By combining equations (15) and (17) and applying the Bayes rule that

$$1/\hat{\omega}_{h,t}^2 = 1/\sigma_h^2 + 1/\hat{\eta}_{h,t}^2, \tag{18}$$

we obtain that

$$\hat{\eta}_{h,t}^2 = \sigma_h^2 / \left(e^{\lambda_{h,t}\theta_h\phi\kappa} - 1 \right). \tag{19}$$

This perceived variance is smaller than the actual variance,

$$\eta_{h,t}^2 = \sigma_h^2 / \left(e^{\lambda_{h,t}\theta_h\kappa} - 1 \right), \tag{20}$$

which we obtain by combining equations (16) and (17) and applying the Bayes rule. Thus, although we model overconfidence as the overestimation of one's information processing ability, our approach is consistent with earlier studies in the finance literature.

Given the extracted signal and its perceived informativeness, the investor updates her belief of the market factor h_{t+1} based on the Bayes rule. The variance of her belief is

$$\hat{\omega}_{h,t}^2 = \sigma_h^2 e^{-\lambda_{h,t}\theta_h\phi\kappa}, \tag{21}$$

and the mean of her belief is

$$\hat{h}_t = \bar{h} + (1 - e^{-\lambda_{h,t}\theta_h\phi\kappa})(s_{h,t} - \bar{h}). \tag{22}$$

We note that $(1 - e^{-\lambda_{h,t}\theta_h\phi\kappa})$, the responsiveness of $\hat{h}_t$ to the signal, increases with ϕ. Therefore, overconfidence leads to an overreaction to the signal.

The investor also overestimates the information content of her signals for the sector and firm-specific factors. We let $\hat{I}(f_{i,t+1}; s_{f,i,t})$, the perceived information amount in the signal about a sector factor $f_{i,t+1}$, be

$$\hat{I}(f_{i,t+1}; s_{f,i,t}) = \frac{1}{2}\phi\theta_i\kappa\lambda_{f,i,t}, \quad i = 1, \cdots, m, \tag{23}$$

where $\kappa\lambda_{f,i,t}$ is the amount of attention the investor allocates to the factor, and $\theta_i > 0$ is the information-efficiency parameter of sector i.

A higher value of θ_i implies that the investor is able to process more information with a given amount of attention. For example, a higher θ_i might correspond to greater analyst coverage, a more transparent financial system and better disclosure, or less complicated fundamentals in sector i. θ_i could be different from θ_h, the information-efficiency parameter of the market factor, and is most likely smaller.

We let $\hat{I}(g_{i,j,t+1}; s_{g,i,j,t})$, the investor's perceived information amount in the signal for a firm-specific factor $g_{i,j,t+1}$, be

$$\hat{I}(g_{i,j,t+1}; s_{g,i,j,t}) = \frac{1}{2}\phi\theta_i\kappa\lambda_{g,i,j,t}, \quad i = 1, \cdots, m, \; j = 1, \cdots, n, \tag{24}$$

where $\kappa\lambda_{g,i,j,t}$ is the amount of attention allocated to the factor, and θ_i is the information-efficiency parameter of sector i. To focus on the difference across various sectors, we let firm-specific factors share the same information-efficiency parameter as their sectors. This assumption is reasonable, since the efficiency of processing information about a firm is positively related to that of the firm's sector.

For the sector and firm-specific factors, we can derive the investor's belief-updating process by following a procedure similar to the one used for the market factor. We denote the investor's posterior belief of a sector factor by

$$f_{i,t+1}|s_{f,i,t} \sim N(\hat{f}_{i,t}, \hat{\omega}_{f,i,t}^2), \quad i = 1, 2, \cdots, m, \tag{25}$$

where the mean and variance are given by

$$\hat{f}_{i,t} = \bar{f} + (1 - e^{-\lambda_{f,i,t}\theta_i\phi\kappa})(s_{f,i,t} - \bar{f}), \tag{26}$$

$$\hat{\omega}_{f,i,t}^2 = \sigma_f^2 e^{-\lambda_{f,i,t}\theta_i\phi\kappa}. \tag{27}$$

The investor's posterior belief in a firm-specific factor is

$$g_{i,j,t+1}|s_{g,i,j,t} \sim N(\hat{g}_{i,j,t}, \hat{\omega}_{g,i,j,t}^2), \quad i = 1, \cdots, m, \ j = 1, \cdots, n, \tag{28}$$

where the mean and variance are given by

$$\hat{g}_{i,j,t} = \bar{g} + (1 - e^{-\lambda_{g,i,j,t}\theta_i\phi\kappa})(s_{i,t} - \bar{g}), \tag{29}$$

$$\hat{\omega}_{g,i,j,t}^2 = \sigma_g^2 e^{-\lambda_{g,i,j,t}\theta_i\phi\kappa}. \tag{30}$$

3.3 The equilibrium

We analyze an equilibrium in which the representative investor holds the net supplies of risky assets according to the market clearing condition. To make an efficient intertemporal consumption decision, the investor allocates her limited attention to analyzing the fundamental factors that affect her portfolio. Following Lucas (1978), we determine the equilibrium asset prices as the shadow prices from the pricing kernel determined by the investor's marginal utility of consumption.

Since the asset fundamentals are independent across periods (no persistence) and the investor has constant risk aversion, the investor's cash holdings would be the only state variable in her optimization problem. In period t, we rewrite her optimization problem in equation (2) as a recursive Bellman equation:

$$J_t(K_t) = \max_{\Lambda_t} E\{\max_{c_t} u(c_t) + \delta E_t[J_{t+1}(K_{t+1})|S_t]\}, \tag{31}$$

12

where J is the investor's value function, K_t is her cash holdings, and S_t is the vector of signals processed by the investor in period t.

Equation (31) illustrates the investor's two optimization decisions: first, she needs to determine her attention allocation Λ_t in information processing, subject to the attention constraint in equation (9); second, based on the processed information, she needs to determine her current period consumption c_t, which trades off her next-period wealth. The investor's next-period cash holding, K_{t+1}, is determined by her current cash holdings, her consumption choice, and the dividends from her risky assets:

$$K_{t+1} = (K_t - c_t)(1 + r) + \sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j,t+1}. \tag{32}$$

We denote the stock price of firm j in sector i at period t by $p_{i,j,t}$, and summarize the prices by a vector:

$$P_{mn \times 1}(t) = (p_{1,1,t}, \cdots, p_{1,n,t}, \cdots, p_{i,j,t}, \cdots, p_{m,1,t}, \cdots, p_{m,n,t})^T. \tag{33}$$

The asset prices are determined by the investor's marginal utility of consumption:

$$P_t = E_t \left[\delta \frac{u'(c_{t+1})}{u'(c_t)} (P_{t+1} + D_{t+1}) \right]. \tag{34}$$

We derive the investor's optimization problem, solve the equilibrium price functions, and verify the equilibrium conditions in Appendix B. The main results of this can be summarized with the following theorem.

Theorem 1 *The investor's attention allocation problem is as follows:*

$$V_t = \min_{\Lambda_t} \ Var_t \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j,t+1} \bigg| S_t \right). \tag{35}$$

The equilibrium asset prices are given by

$$P_t = \frac{1}{1+r} [E(D_{t+1}|S_t) + L_t], \tag{36}$$

where L_t is a deterministic vector.

13

Theorem 1 states that the objective of the investor's attention allocation is to minimize the variance of her belief about the next-period dividend payments to her portfolio. Minimizing future uncertainty allows the investor to make more efficient intertemporal consumption choices. The objective of the investor's attention allocation is independent of the level of asset fundamentals, because these fundamentals are independent over time. In addition, Theorem 1 shows that asset prices are determined by the investor's conditional expectation of the next period's dividends, and therefore fluctuate with the processed information.

Given the linear factor structure in equation (4) and the variances of the investor's conditional beliefs in equations (19), (27) and (30), we transform the investor's attention allocation problem in equation (35) into

$$V_t = \min_{\Lambda_t} \ m^2 n^2 \sigma_h^2 \mathrm{e}^{-\lambda_{h,t}\theta_h\phi\kappa} + \sum_{i=1}^{m} n^2 \sigma_f^2 \mathrm{e}^{-\lambda_{f,i,t}\theta_i\phi\kappa} + \sum_{i=1}^{m}\sum_{j=1}^{n} \sigma_g^2 \mathrm{e}^{-\lambda_{g,i,j,t}\theta_i\phi\kappa} \tag{37}$$

$$\text{subject to} \quad \lambda_{h,t} + \sum_{i=1}^{m} \lambda_{f,i,t} + \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{g,i,j,t} \leq 1, \tag{38}$$

$$\lambda_{h,t} \geq 0, \ \lambda_{f,i,t} \geq 0, \ \lambda_{g,i,j,t} \geq 0, \quad \forall \ i \in \{1,\cdots,m\}, \ j \in \{1,\cdots,n\}. \tag{39}$$

Inequality (38) imposes a constraint on the aggregate attention allocation, and inequalities in (39) require that attention allocation to any factor cannot be negative. The following theorem establishes the existence of a unique solution to this optimization problem. The proof appears in Appendix *C.1*.

Theorem 2 *There exists a unique solution to the investor's attention allocation problem.*

We solve the optimization problem using the Lagrange method:

$$L = m^2 n^2 \sigma_h^2 \mathrm{e}^{-\lambda_{h,t}\theta_h\phi\kappa} + \sum_{i=1}^{m} n^2 \sigma_f^2 \mathrm{e}^{-\lambda_{f,i,t}\theta_i\phi\kappa} + \sum_{i=1}^{m}\sum_{j=1}^{n} \sigma_g^2 \mathrm{e}^{-\lambda_{g,i,j,t}\theta_i\phi\kappa} \tag{40}$$

$$-\mu_{\kappa,t}\left(1 - \lambda_{h,t} - \sum_{i=1}^{m}\lambda_{f,i,t} - \sum_{i=1}^{m}\sum_{j=1}^{n}\lambda_{g,i,j,t}\right) - \mu_{h,t}\lambda_{h,t} - \sum_{i=1}^{m}\mu_{f,i,t}\lambda_{f,i,t} - \sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{g,i,j,t}\lambda_{g,i,j,t}$$

where $\mu_{\kappa,t}$ is the Lagrange multiplier associated with the net allocation constraint in (38), and $\mu_{h,t}$, $\mu_{f,i,t}$ and $\mu_{g,i,j,t}$ are the Lagrange multipliers associated with the non-negativity constraints in (39) on the allocation to the market, sector and firm-specific factors, respectively.

We obtain the following first-order conditions:

$$
\begin{cases}
\mu_{\kappa,t} = \theta_h \phi \kappa (m^2 n^2 \sigma_h^2) \mathrm{e}^{-\lambda_{h,t}\theta_h \phi \kappa} & \text{if} \quad \lambda_{h,t} > 0 \\
\mu_{\kappa,t} > \theta_h \phi \kappa (m^2 n^2 \sigma_h^2) & \text{if} \quad \lambda_{h,t} = 0
\end{cases}
\tag{41}
$$

$$
\begin{cases}
\mu_{\kappa,t} = \theta_i \phi \kappa (n^2 \sigma_f^2) \mathrm{e}^{-\lambda_{f,i,t}\theta_i \phi \kappa} & \text{if} \quad \lambda_{f,i,t} > 0 \\
\mu_{\kappa,t} > \theta_i \phi \kappa (n^2 \sigma_f^2) & \text{if} \quad \lambda_{f,i,t} = 0
\end{cases}
\quad \forall\, i \in \{1, \cdots, m\}
\tag{42}
$$

$$
\begin{cases}
\mu_{\kappa,t} = \theta_i \phi \kappa \sigma_g^2 \mathrm{e}^{-\lambda_{g,i,j,t}\theta_i \phi \kappa} & \text{if} \quad \lambda_{g,i,j,t} > 0 \\
\mu_{\kappa,t} > \theta_i \phi \kappa \sigma_g^2 & \text{if} \quad \lambda_{g,i,j,t} = 0
\end{cases}
\quad \forall\, i \in \{1, \cdots, m\},\ j \in \{1, \cdots, n\}.
\tag{43}
$$

The multiplier $\mu_{\kappa,t}$ represents the opportunity cost of allocating a unit of attention in period t, i.e., more attention to one factor decreases the attention available for processing information about other factors. The first-order conditions in equations (41)-(43) show that a factor will receive a positive amount of attention allocation if and only if its marginal benefit can justify the opportunity cost.

3.4 Comments on the model

One might argue that an attention-constrained investor can choose to "free ride" upon the information revealed through prices, rather than processing the information on her own. However, learning from prices is not exactly free. It still requires attention and knowledge. This learning channel requires knowledge of the market structure, including the presence and risk preference of other traders, their information quality, and the intensity of liquidity trades (Grossman and Stiglitz, 1980). These requirements for using price-based information make this method especially challenging for a retail investor, who typically lacks the knowledge of other market participants' trading motives and cannot observe others' trades in a timely fashion. In addition to these technical difficulties, overconfidence can cause an investor to ignore potentially useful information embedded in prices and to rely only on her own analysis. For simplicity, we restrict the investor in our model from learning through prices.

We also emphasize that in reality retail investors are often exposed to similar information sources, such as media coverage and analysts reports. Thus, even when investors process information independently, their beliefs are likely to be influenced by noise in the common

information sources, causing the errors in their beliefs to be correlated. This correlation prevents the cancellation of investors' learning errors in aggregation.[6] To simplify our exposition, we analyze the learning process of a representative retail investor. Our main results would be viable as long as different investors share common information sources.

We analyze an equilibrium in which the representative investor holds the net supply of the risky assets. It is possible that in a more general model with heterogeneous investors, an individual investor might choose to hold a subset of assets so that she can concentrate her limited attention on a smaller number of uncertain factors. By doing so, she trades off portfolio diversification for learning efficiency. In this paper, we use the representative-investor equilibrium to focus on the effects of investors' category-learning behavior, which we believe is one of the key implications of limited investor attention, and is important for financial markets. The interaction among attention-constrained investors is an interesting issue in its own right. Due to the technical difficulty of incorporating this feature into our current model, we will leave it for future research.

We do not explicitly analyze the reasons why professional investors fail to eliminate the price effects generated by the attention constraints and behavioral biases of individual investors. Instead, we rely on results from earlier papers, e.g., Shleifer and Vishny (1997), which conclude that deviations from efficient prices can be persistent in markets because arbitrage is limited. Arbitrageurs face great risk when trading against unpredictable investor sentiment. At least in the short run, these biased reactions can become even more extreme, moving stock prices further away from their fundamental values and causing losses to arbitrage positions. Since there have been extensive theoretical studies on the limits of arbitrage[7] and because the empirical studies mentioned earlier also confirm that investors' inattention to public information could affect stock prices, we opt to focus on the implications that arise from studying the learning processes of attention-constrained, overconfident investors.

---

[6]Recent empirical studies, e.g., Barber, Odean and Zhu (2003) and Kumar and Lee (2003), confirm that the buy-sell activities of retail investors are correlated. However, Kaniel, Saar, and Titman (2004) noted that the investor sentiment across stocks have low correlations.

[7]Barberis and Thaler (2003) provide a review of these studies.

4 Category learning behavior

Limited investor attention can lead to category learning behavior, i.e., the investor tends to focus more on the market and sector information and less on firm-specific information. Here, we compare the investor's attention allocation for common factors and firm-specific factors.

Here we consider the case in which the information-efficiency parameters for all factors are identical and stay constant over time. The investor's attention allocation is therefore also constant over time. In addition, since all the sector factors are symmetric, with the same prior variances and the same information-efficiency parameters, they receive identical attention allocations. All firm-specific factors also receive identical attention allocations. The following proposition compares the allocations given to different factors. Appendix $C.2$ provides the proof.

Proposition 1 *If $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$, then the investor allocates more attention to the market factor than to a sector factor, and more attention to a sector factor than to a firm-specific factor. When the investor is severely attention-constrained, i.e., $\kappa < \frac{m}{\phi\theta}\log\left(n^2\sigma_f^2/\sigma_g^2\right)$, she devotes attention only to the market and sector factors and pays no attention to firm-specific factors.*

Proposition 1 confirms that the investor tends to allocate more attention to common factors, since it is more effective to concentrate on factors that generate more uncertainty in the portfolio.[8] The conditions that make it possible for this situation to occur, $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$, are realistic. Since there are usually a large number of sectors in the economy and a large number of firms in a sector ($n \gg 1$ and $m \gg 1$), these conditions would hold easily when the uncertainty in the common factors and firm-specific factors have comparable magnitudes. In the extreme case when the investor's attention is severely constrained, he would focus exclusively on common factors and ignore all firm-specific information. Essentially, the investor treats all the stocks in a sector as a single category and only processes information at the category level. He then makes all pricing and investment decisions based purely on this information.

---

[8]Overconfidence also affects the investor's attention allocation. However, since overconfidence makes the investor overestimate both the opportunity cost and the marginal productivity of her attention, the net effect is secondary relative to the effect of uncertainty across factors.

The investor's attention allocation in the severely constrained case dovetails with an observation made by Barberis and Shleifer (2003), that investors often organize their investment decisions based on "style strategies", i.e., they categorize securities into exogenous style classes and shift investment in and out of these classes simultaneously. By analyzing the composition of information processed by an attention-constrained investor, our model allows a more general form of category behavior, i.e., the investor tends to focus more on category-level information than on firm-specific information, and the degree of categorization depends on her attention constraints. This generalization allows us to draw explicit links between the dynamics of information flow and the dynamics of both return comovement and predictability.

We base the categories in our model on asset fundamentals. Although categories are sometimes created through non-fundamental factors, those that are based on fundamentals, such as industries and countries, are certainly important in practice. Cooper, Dimitrov, and Rau (2001) show that during the period 1998-1999, more than 100 companies tried to associate themselves with the high-flying internet industry by changing their company names to dot.com names. Stocks for these companies earned an average abnormal return of 74%, a figure that includes those firms that only changed their names but did not adopt any specific internet related investment or strategies. This example demonstrates that to a great extent, internet stocks during the period were evaluated based solely on their investors' beliefs of the industry as a whole. It also demonstrates how inattentive investors can be to firm-specific information.

5   Implications for return comovement

Information flow is widely regarded as an important source of asset return fluctuation, e.g., French and Roll (1986). Roll (1988) further links the comovement of asset returns to the composition of market and firm-specific information. As more information flows into the market, it seems intuitive that asset prices will become more informative about future fundamentals while asset returns will become more volatile. Furthermore, if there is more market information relative to firm-specific information, asset returns should be more correlated.

5.1 Information flow and return variability

We first clarify the relation between the amount of information incorporated into prices and the corresponding return variability. Here, we take the amount of information as given and examine its impact on return variance.

We consider the excess return of firm j in sector i:

$$q_{i,j,t} = p_{i,j,t} + d_{i,j,t} - (1+r)p_{i,j,t-1}. \tag{44}$$

By substituting the linear price function in Theorem 1 and the factor structure in equation (4) into $q_{i,j,t}$, we can decompose its return variance into three components:

$$Var(q_{i,j,t}) = Var\left(\frac{1}{1+r}\hat{h}_t + h_t - \hat{h}_{t-1}\right) + Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1}\right)$$
$$+ Var\left(\frac{1}{1+r}\hat{g}_{i,j,t} + g_{i,j,t} - \hat{g}_{i,j,t-1}\right). \tag{45}$$

These components respectively correspond to the three fundamental factors: market, sector, and firm-specific.

As an illustration, we consider the return variance component related to a sector factor f_i. Our discussion in this subsection applies equally to all the other factors. There are two parts in the return fluctuation, $Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1}\right)$. The first part, $\frac{1}{1+r}\hat{f}_{i,t}$, represents the fluctuation in price as the investor updates her belief $(\hat{f}_{i,t})$ in response to information on $f_{i,t+1}$, the value of the factor in the following period. The second part, $f_{i,t} - \hat{f}_{i,t-1}$, represents the resolution of the remaining uncertainty in the value of $f_{i,t}$ through the realization of the current-period dividend. The assumption that fundamental factors are independent across periods implies that these two parts are independent. Thus, we can further decompose the return variance component generated by f_i as

$$Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1}\right) = \frac{1}{(1+r)^2}Var(\hat{f}_{i,t}) + Var(f_{i,t} - \hat{f}_{i,t-1}). \tag{46}$$

The amount of information about f_i affects the two return variance components in equation (46) differently. The joint effect depends on the investor's overconfidence, and is summarized in Proposition 2 below. The proof appears in Appendix *C.3*.

When the investor responds to information correctly, as more information is processed, more uncertainty is resolved and the investor's belief updating increases the first part of

return variance, $\frac{1}{(1+r)^2}Var(\hat{f}_{i,t})$. On the other hand, more information processed decreases the remaining uncertainty, thus the other return variance part, $Var(f_{i,t} - \hat{f}_{i,t-1})$, becomes smaller. When the discount rate is zero, more information changes the timing of uncertainty resolution without affecting the return variance. When the discount rate is positive, more information actually decreases the return variance due to the discount factor $\frac{1}{(1+r)^2}$ in the first part of the return variance. This result is given in Part A of Proposition 2, and it confirms the intuition of Campbell et al. (2001) that "improved information about future cash flows increases the volatility of the stock-price level, but it reduces the volatility of the stock return because news arrives earlier, at a time when the cash flows in question are more heavily discounted." Also, we note that this result is not unique to our model specification. West (1988) provides similar results with a more general ARMA-type dividend process.

It seems possible to use a static model with two periods to discuss the relation between information and return variation. For example, in the first period, investors incorporate information to update their beliefs, and determine the equilibrium asset prices; in the second period assets are liquidated. In such a static model, the return variation in the first period reflects only uncertainty resolution through investors' belief updating, while the return variation in the second period reflects only the resolution of the remaining fundamental uncertainty through liquidation. Since both parts of uncertainty resolution are important for understanding return variation, we analyze a dynamic model to avoid any potential confusion. Such a dynamic setup also allows us to discuss the time-series properties of return comovement.

Proposition 2 *The relation between information and return variability depends on the level of the investor's overconfidence.*

A. When the investor uses perfect Bayesian updating ($\phi = 1$), the return variance generated by a fundamental factor is independent of the amount of the information about the factor if the discount rate for future cashflows is zero, and decreases with the amount of information if the discount rate is positive.

B. If the investor is sufficiently overconfident:

$$\phi > 1 + \frac{(1+r)^2 - 1}{(1+r)^2 + 1}, \tag{47}$$

with the amount of information below a threshold, the return variance increases with the amount of information and the overconfidence parameter ϕ.

Part B of Proposition 2 shows that when the investor is overconfident, more information can lead to a higher return variability. Overconfidence causes the investor to overreact to the self-processed noisy signal, thus amplifying the effect of signal noise on prices. The exact magnitude of the noise amplification effect depends on two forces. On the one hand, as a signal becomes more precise, the investors' overreaction to the signal gets stronger, thus further amplifying noise. On the other hand, a more precise signal contains less noise. The tradeoff between these two forces determines that when the investor's overconfidence parameter is sufficiently high and the amount of information in the signal is below the threshold level, the noise amplification effect can dominate the discount rate effect that we mentioned earlier. Thus, in the presence of investor overconfidence, more information can lead to a higher return variance. We combine this result with the investor's attention allocation to analyze return comovement.

In Appendix D, we numerically illustrate values of the threshold on the amount of information, based on various parameter sets. These threshold values are high for individual investors' learning processes. We also show that the restriction on the overconfidence parameter in (47) is modest for reasonable values of the discount rate. Therefore, we focus the rest of our analysis on cases in which the conditions in Proposition 2B are satisfied.

5.2 "Excess" comovement

The return correlation between two firms depends on the fraction of their return variances that is driven by common sources. Here, we compare the correlations in firms' returns with the correlations in their fundamentals.

We define ρ as the return correlation between two firms from different sectors:

$$\rho \equiv Corr(q_{i,j,t}, q_{k,l,t}), \quad i \neq k. \tag{48}$$

Since these two firms are from different sectors, according to equation (45) the shared part

of their return variances can only be related to the market factor:

$$\rho = \frac{Var\left(\frac{1}{1+r}\hat{h}_t + h_t - \hat{h}_{t-1}\right)}{\sqrt{Var(q_{i,j,t})Var(q_{k,l,t})}}. \tag{49}$$

We define ρ_i as the return correlation between any two firms in sector i:

$$\rho_i \equiv Corr(q_{i,j,t},\ q_{i,k,t}), \qquad j \neq k. \tag{50}$$

According to equation (45), ρ_i depends on the return variances caused by the market factor and the sector factor:

$$\rho_i = \frac{Var\left(\frac{1}{1+r}\hat{h}_t + h_t - \hat{h}_{t-1}\right) + Var\left(\frac{\hat{f}_{i,t}}{1+r} + f_{i,t} - \hat{f}_{i,t-1}\right)}{\sqrt{Var(q_{i,j,t})Var(q_{i,k,t})}}. \tag{51}$$

The following proposition compares the return correlations with the corresponding fundamental correlations. The proof appears in Appendix *C.4*.

Proposition 3 *Under the conditions in Proposition 2B, and that $m^2\sigma_h^2 > \sigma_f^2$, $n^2\sigma_f^2 > \sigma_g^2$, the return correlation between any two firms is higher than their fundamental correlation.*

Proposition 3 shows that limited attention and investor overconfidence can amplify the return correlation between any two firms, either in the same sector or from different sectors, so that the return correlation is higher than their fundamental correlation. This result is a consequence of the investor's category learning behavior. His tendency to process more market than sector information and to process more sector than firm-specific information, together with his overreaction to this processed information, creates more return variation in common factors and therefore generates excessive comovement.

The result of excess comovement at the market level is supported by empirical evidence of Shiller (1989) and Pindyck and Rotemberg (1993). Shiller finds that the comovement of real stock prices between the U.K. and the U.S. appear too large to be accounted for in terms of the comovement of real dividends between the two countries. Pindyck and Rotemberg show a similar result for the comovement of individual stocks in U.S. markets. The result of excess comovement at the sector level also complements the analysis of Barberis and Shleifer (2003)

and Barberis, Shleifer, and Wurgler (2005), who focus on excess comovement generated by investors' style strategies among assets with no fundamental correlation.[9]

5.3 Comovement and price informativeness

The degree to which the investor treats firms in a sector as a single category depends on the information-efficiency parameter of the sector. We compare cross-sectoral differences in the information-efficiency parameters with a focus on return correlation and price informativeness.

We capture a firm's stock price informativeness through its inverse relationship to remaining uncertainty, which is to be resolved through the realization of dividends:

$$Var(h_t - \hat{h}_{t-1}) + Var(f_{i,t} - \hat{f}_{i,t-1}) + Var(g_{i,j,t} - \hat{g}_{i,j,t-1}). \tag{52}$$

The firm's stock price informativeness depends on the amount of information about the firm that the investor processes. For a given overconfidence parameter, Lemma 3 in Appendix *C.3* shows that as the investor processes more information, the stock price becomes more informative and the remaining uncertainty is lower.

The following proposition compares the return correlation and price informativeness in two sectors with different information-efficiency parameters. The proof is in Appendix *C.5*.

Proposition 4 *Assume that $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$. Consider two sectors, 1 and 2, with different information-efficiency parameters, $\theta_1 > \theta_2$. In sector 1, the investor not only processes more sector and firm-specific information, but also processes relatively more firm-specific information, treating firms in this sector less as a single category and more as individual firms. If the conditions in Proposition 2B are satisfied and all factors receive positive attention allocation, stock prices in sector 1 are more informative and the return correlation between firms in sector 1 is lower.*

Proposition 4 illustrates a negative cross-sectoral relation between return comovement and stock price informativeness. Sector 1 has a higher information-efficiency parameter, e.g., either a better disclosure system or a less complicated industry structure. As a result,

---

[9]See also Boyer (2004) for a study of excessive comovement among stocks with similar book-to-market ratios.

it is effective for the investor to allocate more attention to all factors in sector 1, leading to more informative prices in the sector. In this scenario, she would also allocate relatively more attention to firm-specific factors in sector 1. This allows her to consider firms within this higher efficiency sector individually rather than as members of a sector category, resulting in lower average return correlation.

We can quantify price informativeness empirically by the power of stock returns in predicting future changes in firms' earnings or other fundamental variables.[10] We can also measure the return correlation between firms in a sector by the average R^2 statistic of regressing individual stock returns in the sector onto the market return and the sector's aggregate return.

Proposition 4 is consistent with several recent empirical studies that examine the relation between comovement and stock price informativeness. According to Morck, Yeung, and Yu (2000), return synchronicity, as measured by the average R^2 statistic, is high in those emerging countries having less developed legal and investor protection systems that discourage information processing. Durnev et al. (2003) show that in industries with less synchronous returns, stock prices are more informative. Wurgler (2000) finds that the efficiency of capital allocation across countries is negatively correlated with the synchronicity in stock returns of domestically traded firms, and Durnev, Morck, and Yeung (2004) show that U.S. firms with higher firm-specific return variability make more value-enhancing investments. Overall, these studies find broad evidence that a sector's stock price informativeness is negatively related to the return correlation of firms in the sector.

Proposition 4 also suggests that the degree to which the investor treats firms in a sector as a single category is determined by her attention allocation to the sector factor relative to firm-specific factors, and that this degree is inversely related to the average return correlation of firms in the sector. Although it is difficult to directly measure the investor's attention allocations, we can use the average return correlation as an indirect measure. This is a useful gauge for analyzing cross-sectional implications of investors' attention.

---

[10]Durnev, et al. (2003) provide a detailed discussion of the empirical measures.

5.4 Time trend in return correlation

Over the last 50 years, enormous developments in information technology have made it easier for investors to collect and process information about the overall economy and to gain specific knowledge about industries and firms. We capture these technological advances by allowing an increasing time trend in the information-efficiency parameter. We assume that all factors share the same information-efficiency parameter θ, which increases over time. The following proposition describes the result. The proof appears in Appendix $C.6$.

Proposition 5 *Suppose that $m^2\sigma_h^2 > \sigma_f^2$, $n^2\sigma_f^2 > \sigma_g^2$, and that the conditions of Proposition 2B are satisfied. As θ increases over time, the investor allocates relatively more attention to firm-specific factors. Consequently, the firm-specific return volatility rises relative to the sector-level volatility, and the return correlation between any two firms decreases.*

Proposition 5 indicates that as information technology improves over time, the investor's attention constraints become less binding, giving her the capacity to process more information for all factors. There is a larger increase in the amount of information she processes for firm-specific factors, and therefore a larger increase in assets' return variances related to firm-specific information. As a result, firm-specific volatility rises and return correlation decreases.

This implication is supported by Campbell, Lettau, Malkiel, and Xu (2001), who find a decreasing trend in the return correlation of U.S. stocks over the last 30 years: the average correlation among individual stocks has fallen from 0.28 in the early 1960s to 0.08 in 1997, and the average R^2 statistics for individual stocks from a market model has declined in tandem. They identify increased firm-specific volatility as the main cause of the decreasing correlation in U.S. stock returns.

We note that the variation in information acquisition cost alone is not sufficient to explain either the cross-sectional or the time-serial properties of return comovement that our model generates. As shown earlier, more information in a rational expectations model with a positive discount rate reduces, rather than increases, return volatility. Without exploring investor behavior such as overconfidence and limited attention, standard rational expectations models cannot reconcile these aforementioned empirical observations.

6 Cross-section of return predictability

Over the past two decades, researchers have discovered that they can use past stock prices and other publicly available variables to predict stock returns.[11] These findings have challenged standard asset pricing models and motivated studies on the effects of investor biases.

We focus on two types of information related biases: investors' overreaction to their information, and their ignorance of useful public information. Investors' attention allocation could play important roles for both types of biases: biased reactions to information only occur when investors attend to information; the predictive power of a piece of ignored information depends on how much the investor already knows. Thus, the investor's attention allocation in our model provides cross-sectional implications for both types of return predictability.

6.1 Predictability by overreactions to information

Overconfidence causes the investor to overreact to her information, causing asset prices to deviate from their fundamental levels. Since the overreactions will eventually be corrected and asset prices will mean-revert to the fundamental levels in the long run, asset returns become predictable. Such predictability is displayed through a negative autocovariance in returns.[12]

Proposition 6 summarizes our results on overreaction driven predictability. We provide the proof in Appendix *C.7*.

Proposition 6 *If the investor is overconfident ($\phi > 1$), the first-order autocovariance of any firm's stock return is negative. Under the condition of Proposition 2B, the autocovariance in the firm-specific returns decreases with the amount of firm-specific information processed. If the investor only attends to information on common factors, the negative autocovariance is present in the common return components, but not in the firm-specific returns.*

Proposition 6 shows that when the investor treats the sector more as a category in itself and processes less firm-specific information, there is less overreaction-driven return

---

[11]For example, the literature has identified the so-called "short-term momentum effect", "long-term reversal effect", "post earnings announcement drift effect" and "accrual effect", among others. See Hirshleifer (2001) and Barberis and Thaler (2003) for reviews of these studies.

[12]Lo and MacKinlay (1988) and Poterba and Summers (1988) find that U.S. stock returns over one year or longer horizons have a negative autocovariance.

predictability at the firm level. In extreme cases where the investor treats all firms in the sector identically as a category and pays no attention to firm-specific information, negative autocovariances are present only in the common return components, such as market- and sector-related returns. Even though stock prices in these cases could be inefficient, in that they do not reflect firm-specific fundamentals, there is no overreaction-driven predictability in the sector's firm-specific returns.

Empirically, it is difficult to directly measure the degree to which the investor treats firms in a sector as a category. However, our analysis in Section 5.3 shows that such a degree is negatively related to the average return correlation of firms in the sector. Thus, our model provides a new testable implication for overreaction-driven return predictability and return comovement: we expect that after controlling for the degree of investor overconfidence, a sector with a lower average return correlation to exhibit more pronounced firm-level return predictability driven by investors' overreaction, such as long-run reversals in asset prices.

Although our model only incorporates return predictability generated by investor over-confidence, we expect the cross-sectional patterns driven by investors' attention allocation to be applicable to other information-related biases. For example, Daniel, Hirshleifer, and Subrahmanyam (1998) show that self-attribution bias can lead to price momentum. Self-attribution bias causes investors to attribute success to themselves but failure to external forces. Thus, investors become more overconfident about their information-processing ability after good performance, and react even more strongly to their information, creating price momentum. Given investors' attention allocation and the resulting information structure, an extension of our model incorporating self-attribution bias could generate a negative relation between momentum in firm-specific returns and average return correlation in a sector.

6.2   Predictive power of ignored information

Limited attention also causes the investor to ignore certain useful information in the public domain. As a result, it is possible that this information can be used to predict future asset returns. Here, we expand our model to discuss the cross-sectoral pattern in the return predictive power of ignored information.

We consider a specific public signal, $\nu_{i,j,t}$, that is ignored by the investor at time t:

$$\nu_{i,j,t} = g_{i,j,t+1} + \epsilon_{\nu,i,j,t} \tag{53}$$

where $g_{i,j,t+1}$ is a firm's fundamental factor and $\epsilon_{\nu,i,j,t} \sim N(0, \sigma_\nu^2)$ is an independent signal noise. This signal could be the cashflow component of the firm's earnings, which is often ignored by investors and thus has the ability to predict future stock returns (see, for example, Sloan, 1996 and Hirshleifer et al., 2003).

The information value of an ignored signal depends on how much information the investor has already processed from other sources, which in turn is determined by the attention she allocates to the firm. Take the EntreMed example (Huberman and Regev, 2001): while investors may have missed the article in *Nature* about the company's new cancer curing drug, they could have obtained this same information if they had paid enough attention to the company information available through other channels. The following proposition explicitly describes such a relationship. We provide the proof in Appendix *C.8*.

Proposition 7 *The return predictive power of an ignored signal about a firm decreases with the amount of firm-specific information processed by the investor.*

Proposition 7 confirms that an investor's allocation of more attention to a firm and processing of more information decreases the remaining uncertainty, which makes the ignored signal less important. Although it can be difficult to measure the investor's attention to an individual firm, our analysis in Section *5.3* shows that the investor's attention to individual firms in a sector is negatively related to the average return correlation of firms in the sector. Thus, our model provides another testable implication: in a sector with a higher average return correlation, a variable that captures ignored firm information should have a higher predictive power for firm-specific returns, after we control for the precision of the variable and the degree of investor overconfidence.

7   Conclusion

In this paper, we provide a theoretical model for studying investors' attention allocation in learning about asset fundamentals. Our model shows that limited attention generates an

endogenous structure of information that reflects the investor's category learning behavior, i.e., the investor tends to focus more on market- and sector-level information than on firm-specific information. When we combine this endogenous information structure with investor overconfidence, we generate several implications for asset return comovement. We show that return correlation between firms is higher than their fundamental correlation; stock prices in a sector with a lower average return correlation tend to be more informative; return correlations decrease over time with the development of information technology. These results are supported by recent empirical findings, which would otherwise be difficult to explain using standard rational expectations models.

In addition, our model demonstrates new implications for the cross-sectional patterns of return predictability: after controlling for the degree of investor overconfidence, firms in a sector with a lower average return correlation tend to have more pronounced overreaction-driven return predictability, such as long-run price reversals and short-term price momentum; on the other hand, ignored information in public domain, such as certain variables in firms' financial statements, is less effective in predicting the future returns of these firms.

Appendix A The entropy measure of information

This appendix provides a brief introduction to the entropy concept. According to information theory, the uncertainty of a random variable $\tilde{y}$ with a continuous probability density function $f(y)$ can be measured by its entropy (Cover and Thomas, 1991, Chapter 2):

$$H(\tilde{y}) \equiv -E\log[f(\tilde{y})] = -\int (\log f(y)) f(y) dy. \tag{A1}$$

Entropy provides a lower bound for the average number of bits needed to transmit the random variable.

The entropy of a random variable that follows a Gaussian distribution, $N(\bar{y}, \sigma^2)$, depends only on its variance:

$$H(\tilde{y}) = \frac{1}{2}\log \sigma^2 + 0.5\log(2\pi e). \tag{A2}$$

This expression is intuitive, since, for a normally distributed random variable, more uncertainty is associated with a larger variance.

Information theory defines the conditional entropy of $\tilde{y}$, given another random variable s, as the expected value of the conditional distributions' entropies, averaged over the conditioning random variable s. Let $f(\tilde{y}, s)$ be the joint probability density function of $\tilde{y}$ and s, $f(y|s)$ be the conditional distribution of $\tilde{y}$ conditional on s, and $g(s)$ be the marginal density function of s. The conditional entropy of $\tilde{y}$ given s is:

$$H(\tilde{y}|s) \equiv \int H(\tilde{y}|s)g(s)ds = -\int\int (\log f(y|s)) f(y,s) dy ds. \tag{A3}$$

The amount of information that s contains about $\tilde{y}$ is the amount of uncertainty reduction in $\tilde{y}$ due to the knowledge of s:

$$I(\tilde{y}; s) \equiv H(\tilde{y}) - H(\tilde{y}|s) = \int\int f(y,s)\log\frac{f(y,s)}{f(y)g(s)} dy ds. \tag{A4}$$

A property of this information measure is that it is invariant to any linear transformations of $\tilde{y}$ and s:

Lemma 1 *If we transform random variables $\tilde{y}$ and s by any non-zero constants a and b, respectively, the amount of information contained in bs about $a\tilde{y}$ is identical to the amount of information contained in s about $\tilde{y}$: $I(a\tilde{y}; bs) = I(\tilde{y}; s)$.*

PROOF: Define $\tilde{y}' = a\tilde{y}$ and $s' = bs$. The joint density function of $\tilde{y}'$ and s' is

$$f_{\tilde{y}'s'}(y', s') = \frac{1}{ab}f(y'/a, s'/b). \tag{A5}$$

The marginal density functions of $\tilde{y}'$ and s' are

$$f_{\tilde{y}'}(y') = \frac{1}{a}f(y'/a), \quad f_{s'}(s') = \frac{1}{b}g(s'/b). \tag{A6}$$

Then the amount of information of bs to $a\tilde{y}$ is given by

$$
\begin{aligned}
I(a\tilde{y}; bs) = I(\tilde{y}'; s') &= \int\int f_{\tilde{y}'s'}(y', s') \log \frac{f_{\tilde{y}'s'}(y', s')}{f_{\tilde{y}'}(y')g_{\tilde{y}'}(s')}dy'ds' \\
&= \int\int \frac{1}{ab}f(y'/a, s'/b) \log \frac{f(y'/a, s'/b)}{f(y'/a)g(s'/b)}dy'ds' \\
&= \int\int f(y, s) \log \frac{f(y, s)}{f(y)g(s)}dyds = I(\tilde{y}; s). \tag{A7}
\end{aligned}
$$

Q.E.D.

This property establishes entropy reduction as a measure of information that is independent of the scale of the underlying variables. Other measures of information have been used in the related literature. For example, Verrecchia (1982) uses the precision (the inverse of variance) of noisy signals with Gaussian distributions to measure information. However, signal precision is not invariant to the scale of the underlying variable to be inferred.

Appendix B   The investor's optimization problem and equilibrium prices

We solve the investor's optimization problem in equation (31) using backward induction. We first derive the investor's consumption decision given her information about the next-period asset fundamentals, and the objective function of the investor's attention allocation. We then solve the equilibrium asset prices from the pricing kernel determined by the investor's marginal utility of consumption.

Given the investor's belief about asset fundamentals, her optimization problem is a standard one, as in Merton (1971). Since the investor has an exponential utility and a Gaussian wealth distribution, her value function takes the following form:

$$J_t(K_t) = -\frac{1}{\gamma}e^{-a_t K_t - b_t}, \tag{B1}$$

where a_t and b_t are deterministic coefficients. Given the investor's signals S_t in period t, she determines her consumption by solving the following optimization problem (inside the curly bracket of equation 31):

$$
\begin{aligned}
J_t(K_t, S_t) &\equiv \max_{c_t} \ u(c_t) + \delta E_t[J_{t+1}(K_{t+1})|S_t] \\
&= \max_{c_t} \ -\frac{1}{\gamma}e^{-\gamma c_t} - \frac{\delta}{\gamma}E_t\left\{ e^{-a_{t+1}\left[(K_t-c_t)(1+r)+\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}\right]-b_{t+1}} \middle| S_t \right\}
\end{aligned} \tag{B2}
$$

where the last equation is derived by substituting in $J_{t+1}(K_{t+1})$ from equation (B1) and K_{t+1} from equation (32). The first-order condition is

$$
u'(c_t) = \delta(1+r)E_t[J'_{t+1}(K_{t+1})|S_t], \tag{B3}
$$

or equivalently,

$$
e^{-\gamma c_t} = \frac{\delta(1+r)a_{t+1}}{\gamma}e^{-(1+r)a_{t+1}(K_t-c_t)-b_{t+1}-a_{t+1}E_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right)+\frac{a_{t+1}^2}{2}Var_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right)} \tag{B4}
$$

From this equation, we can derive that

$$
\begin{aligned}
c_t &= \frac{1}{\gamma+(1+r)a_{t+1}}\left\{ (1+r)a_{t+1}K_t + b_{t+1} + a_{t+1}E_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right) \right. \\
&\left. -\frac{a_{t+1}^2}{2}Var_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right) - \ln[\delta(1+r)a_{t+1}/\gamma] \right\}. \tag{B5}
\end{aligned}
$$

By substituting c_t back to equation (B2), we obtain

$$
\begin{aligned}
&J_t(K_t, S_t) \\
&= -\frac{\gamma+(1+r)a_{t+1}}{\gamma(1+r)a_{t+1}}exp\left\{ -\frac{\gamma}{\gamma+(1+r)a_{t+1}}\left[(1+r)a_{t+1}K_t + a_{t+1}E_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right) \right.\right. \\
&\left.\left. -\frac{a_{t+1}^2}{2}Var_t\left(\sum_{i=1}^{m}\sum_{j=1}^{n}d_{i,j,t+1}|S_t\right) + b_{t+1} - \ln[\delta(1+r)a_{t+1}/\gamma] \right] \right\}. \tag{B6}
\end{aligned}
$$

From equation (31), the investor's attention allocation is determined by solving the following optimization problem:

$$
J_t(K_t) = \max_{\Lambda_t} \ E[J_t(K_t, S_t)]. \tag{B7}
$$

The expectation on the right-hand-side of equation (B7) is based on the investor's prior belief at t. Since the investor determines the attention allocation before she processes any information about the next-period dividends, and because of the i.i.d assumption of dividends across periods, she chooses her attention allocation to maximize the unconditional expectation of $J_t(K_t, S_t)$.

After substituting $J_t(K_t, S_t)$ from equation (B6) into equation (B7) and deriving the expectation, the objective of the investor's attention allocation is to

$$\min_{\Lambda_t} \; Var_t \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j,t+1} | S_t \right). \tag{B8}$$

This term is the only one affected by the investor's attention allocation on the right-hand-side of equation (B7). Intuitively, the objective is to minimize the total uncertainty about the next-period dividend payments in the investor's portfolio. In achieving such an objective, the investor is able to make an efficient intertemporal consumption decision.

The exact values of the coefficients a_t and b_t are not critical for our discussion. They are determined by a recursive Bellman equation. By substituting equations (B1) and (B6) into the two sides of equation (B7), respectively, and by matching the corresponding terms, we obtain two backward-induction equations that determine a_t and b_t:

$$a_t = \frac{\gamma(1+r)a_{t+1}}{\gamma + (1+r)a_{t+1}}, \tag{B9}$$

$$e^{-b_t} = \frac{\gamma + (1+r)a_{t+1}}{(1+r)a_{t+1}} E \left\{ exp \left\{ -\frac{\gamma}{\gamma + (1+r)a_{t+1}} \left[a_{t+1} E_t \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j,t+1} | S_t \right) + b_{t+1} \right. \right. \right.$$
$$\left. \left. \left. - \frac{a_{t+1}^2}{2} Var_t \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j,t+1} | S_t \right) - \ln[\delta(1+r)a_{t+1}/\gamma] \right] \right\} \right\}. \tag{B10}$$

Following Lucas (1978), the investor's marginal utility of consumption determines the asset prices:

$$P_t = E_t \left[\delta \frac{u'(c_{t+1})}{u'(c_t)} (P_{t+1} + D_{t+1}) | S_t \right]. \tag{B11}$$

Since the dividends are independent across periods, we conjecture the following linear price function:

$$P_t = \frac{1}{1+r} [E_t(D_{t+1}|S_t) + L_t] \tag{B12}$$

33

where L_t is a deterministic vector. L_t can be determined through backward induction by substituting equation (B12) into equation (B11):

$$
\begin{aligned}
P_t &= \delta e^{\gamma c_t} E_t \left\{ e^{-\gamma c_{t+1}} \left[\frac{1}{1+r} [E_{t+1}(D_{t+2}|S_{t+1}) + L_{t+1}] + D_{t+1} \right] |S_t \right\} \\
&= \delta e^{\gamma c_t} \{ E_t[e^{-\gamma c_{t+1}}](\bar{D} + L_{t+1})/(1+r) + E_t[e^{-\gamma c_{t+1}} D_{t+1}|S_t] \} \\
&= [(\bar{D} + L_{t+1})/(1+r) + E_t(D_{t+1}) - Cov_t(\gamma c_{t+1}, D_{t+1})] E_t \left[\delta \frac{u'(c_{t+1})}{u'(c_t)} \right]. \quad (B13)
\end{aligned}
$$

The last equation uses Lemma 2, given below. The envelope condition implies that

$$
J_t'(K_t) = u'(c_t). \quad (B14)
$$

Combining this condition with the first-order condition in equation (B3), we obtain

$$
E_t \left[\delta \frac{u'(c_{t+1})}{u'(c_t)} \right] = \frac{1}{1+r}. \quad (B15)
$$

By substituting this expression into equation (B13), we obtain a recursive equation that determines the deterministic vector L_t:

$$
L_t = \frac{1}{1+r}(\bar{D} + L_{t+1}) - Cov_t(\gamma c_{t+1}, D_{t+1}). \quad (B16)
$$

Thus, the conjectured price function in equation (B12) holds in equilibrium.

Lemma 2 *Let x and y be two normally distributed random variables. Then,*

$$
E[ye^x] = [E(y) + Cov(x, y)]E(e^x). \quad (B17)
$$

PROOF: Denote $x \sim N(\bar{x}, \sigma_x^2)$ and $y \sim N(\bar{y}, \sigma_y^2)$, and their correlation as ρ. Then,

$$
\begin{aligned}
E[ye^x] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \frac{ye^x}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\bar{x}}{\sigma_x}\right)\left(\frac{y-\bar{y}}{\sigma_y}\right) + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 \right]} \\
&= \int_{-\infty}^{\infty} dx \frac{e^{x-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^2}}{2\pi \sigma_x} \int_{-\infty}^{\infty} dy \frac{y}{\sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\bar{y}}{\sigma_y} - \rho\frac{x-\bar{x}}{\sigma_x}\right)^2} \\
&= \int_{-\infty}^{\infty} dx \frac{e^{x-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^2}}{2\pi \sigma_x} [\bar{y} + \frac{\rho\sigma_y}{\sigma_x}(x - \bar{x})] = (\bar{y} + \rho\sigma_x\sigma_y)e^{\bar{x}+\sigma_x^2/2}. \quad (B18)
\end{aligned}
$$

Q.E.D.

Appendix C Other proofs

C.1 *Proof to Theorem 2*

Since the objective function to be minimized is continuous and the feasible set of variables is compact, there must exist at least one minima for the optimization problem, as directly implied by the Weierstrass Theorem. On the other hand, since the objective function is strictly convex, the minima must also be unique.

C.2 *Proof to Proposition 1*

Since all sector factors are symmetric with an identical prior variance and information-efficiency parameter, and since these parameters stay constant over time, they receive identical constant attention allocations: $\lambda_{f,1} = \cdots = \lambda_{f,m}$, which we denote by λ_f. By construction, $\lambda_f \leq 1/m$. All firm-specific factors also receive an identical attention allocation, which we denote by λ_g. In addition, we denote the investor's attention allocation to the market factor as λ_h.

We first show that if $n^2\sigma_f^2 > \sigma_g^2$, then $\lambda_f \geq \lambda_g$. We prove this statement by contradiction. We suppose that $0 \leq \lambda_f < \lambda_g$. Due to the stationary setting, we can rewrite the first-order conditions in equations (42) and (43) as

$$
\begin{cases}
\mu_\kappa = \theta\phi\kappa(n^2\sigma_f^2)\mathrm{e}^{-\lambda_f\theta\phi\kappa} & \text{if } \lambda_f > 0 \\
\mu_\kappa > \theta\phi\kappa(n^2\sigma_f^2) & \text{if } \lambda_f = 0
\end{cases}
\tag{C1}
$$

$$
\begin{cases}
\mu_\kappa = \theta\phi\kappa\sigma_g^2\mathrm{e}^{-\lambda_g\theta\phi\kappa} & \text{if } \lambda_g > 0 \\
\mu_\kappa > \theta\phi\kappa\sigma_g^2 & \text{if } \lambda_g = 0
\end{cases}
\tag{C2}
$$

λ_g must be positive. Then the first-order conditions imply that

$$
\mu_\kappa \;\geq\; \theta\phi\kappa(n^2\sigma_f^2)\mathrm{e}^{-\lambda_f\theta\phi\kappa}
\tag{C3}
$$

$$
\mu_\kappa \;=\; \theta\phi\kappa\sigma_g^2\mathrm{e}^{-\lambda_g\phi\kappa}
\tag{C4}
$$

Thus, $n^2\sigma_f^2\mathrm{e}^{-\lambda_f\theta\phi\kappa} \leq \sigma_g^2\mathrm{e}^{-\lambda_g\theta\phi\kappa}$. On the other hand, since $n^2\sigma_f^2 > \sigma_g^2$ and $\lambda_f < \lambda_g$ by assumptions, $n^2\sigma_f^2\mathrm{e}^{-\lambda_f\theta\phi\kappa} > \sigma_g^2\mathrm{e}^{-\lambda_g\theta\phi\kappa}$. This result is a contradiction. Therefore, $\lambda_f \geq \lambda_g$.

By following the logic as above, we can also show that if $m^2\sigma_h^2 > \sigma_f^2$, then $\lambda_h \geq \lambda_f$.

If $\kappa < \frac{m}{\phi\theta} \log\left(n^2\sigma_f^2/\sigma_g^2\right)$, we can further show that λ_g must be zero. First we note that the assumption implies that

$$(n^2\sigma_f^2)\mathrm{e}^{-\theta\phi\kappa/m} > \sigma_g^2. \tag{C5}$$

Starting from the first order conditions in equation (C1), we obtain the following:

$$\mu_\kappa \geq \theta\phi\kappa(n^2\sigma_f^2)\mathrm{e}^{-\lambda_f\theta\phi\kappa} \geq \theta\phi\kappa(n^2\sigma_f^2)\mathrm{e}^{-\theta\phi\kappa/m} > \theta\phi\kappa\sigma_g^2, \tag{C6}$$

where we derive the second inequality above from the fact that $\lambda_f \leq 1/m$, and the third inequality from (C5). The first-order condition in equation (C2) then implies that $\lambda_g = 0$.

C.3 Proof to Proposition 2

Define

$$v_p(x, \phi) \equiv (1 - e^{-\phi x})^2 \frac{1}{1 - e^{-x}}, \tag{C7}$$

$$v_d(x, \phi) \equiv e^{-2\phi x} + (1 - e^{-\phi x})^2 \frac{e^{-x}}{1 - e^{-x}}. \tag{C8}$$

Lemma 3 derives the two parts of return variance in equation (46), $Var(\hat{f}_{i,t})$ and $Var(f_{i,t} - \hat{f}_{i,t-1})$, based on these two functions.

Lemma 3 $Var(\hat{f}_{i,t})$ *is determined by the amount of information processed by the investor,* $\lambda_{f,i,t}\theta_i\kappa$, *about the value of the factor in the next period,* $f_{i,t+1}$,:

$$Var(\hat{f}_{i,t}) = v_p(\lambda_{f,i,t}\theta_i\kappa, \phi)\sigma_f^2. \tag{C9}$$

$Var(f_{i,t} - \hat{f}_{i,t-1})$ *is determined by* $\lambda_{f,i,t-1}\theta_i\kappa$, *the amount of information processed in the previous period about* $f_{i,t}$, *which affects the remaining uncertainty in* $f_{i,t}$:

$$Var(f_{i,t} - \hat{f}_{i,t-1}) = v_d(\lambda_{f,i,t-1}\theta_i\kappa, \phi)\sigma_f^2. \tag{C10}$$

$v_d(x, \phi)$ *is a monotonically decreasing function of* x *for any* $x \geq 0$ *and* $\phi \geq 1$.

PROOF: According to equation (26),

$$Var(\hat{f}_{i,t}) = (1 - e^{-\lambda_{f,i,t}\theta_i\phi\kappa})^2 Var(s_{f,i,t})$$

36

$$
\begin{aligned}
&= (1 - e^{-\lambda_{f,i,t}\theta_i\phi\kappa})^2 [Var(f_{i,t+1}) + Var(\epsilon_{f,i,t})] \\
&= (1 - e^{-\lambda_{f,i,t}\theta_i\phi\kappa})^2 \frac{\sigma_f^2}{1 - e^{-\lambda_{f,i,t}\theta_i\kappa}} \\
&= \upsilon_p(\lambda_{f,i,t}\theta_i\kappa, \phi)\sigma_f^2, \quad\quad\quad\quad\quad\quad\quad\quad\text{(C11)}
\end{aligned}
$$

where the second equation above is obtained through equation (11), and the third equation is derived through equation (20). Similarly,

$$
\begin{aligned}
Var(f_{i,t} - \hat{f}_{i,t-1}) &= Var\{f_{i,t} - [\bar{f} + (1 - e^{-\lambda_{f,i,t-1}\theta_i\phi\kappa})(s_{f,i,t-1} - \bar{f})]\} \\
&= Var\{f_{i,t} - [\bar{f} + (1 - e^{-\lambda_{f,i,t-1}\theta_i\phi\kappa})(f_{i,t} + \epsilon_{f,i,t-1} - \bar{f})]\} \\
&= Var[e^{-\lambda_{f,i,t-1}\theta_i\phi\kappa}(f_{i,t} - \bar{f})] + Var[(1 - e^{-\lambda_{f,i,t-1}\theta_i\phi\kappa})\epsilon_{f,i,t-1}] \\
&= \left[e^{-2\lambda_{f,i,t-1}\theta_i\phi\kappa} + (1 - e^{-\lambda_{f,i,t-1}\theta_i\phi\kappa})^2 \frac{e^{-\lambda_{f,i,t-1}\theta_i\kappa}}{1 - e^{-\lambda_{f,i,t-1}\theta_i\kappa}} \right] \sigma_f^2 \\
&= \upsilon_d(\lambda_{f,i,t-1}\theta_i\kappa, \phi)\sigma_f^2. \quad\quad\quad\quad\quad\quad\text{(C12)}
\end{aligned}
$$

For any $x > 0$,

$$
\begin{aligned}
\frac{\partial \upsilon_d(x, \phi)}{\partial x} &= -\frac{1}{(1 - e^{-x})^2} \Big\{ 2\phi e^{-\phi x}(1 - e^{-x})[e^{-\phi x} - e^{-(\phi+1)x}] \\
&\quad\quad + [1 - (\phi+1)e^{-\phi x} + \phi e^{-(\phi+1)x}][e^{-x} - e^{-(\phi+1)x}] \Big\} < 0. \quad\text{(C13)}
\end{aligned}
$$

Thus, $\upsilon_d(x, \phi)$ always decreases with x. Q.E.D.

We consider a stationary equilibrium in which the amount of information processed in each period about a factor f_i is identical: $\lambda_{f,i,t}\theta_i\kappa = \lambda_{f,i,t-1}\theta_i\kappa = x$. Define

$$
\upsilon(x, \phi) = \frac{1}{(1+r)^2}\upsilon_p(x, \phi) + \upsilon_d(x, \phi). \quad\quad\quad\quad\text{(C14)}
$$

The total return variance generated by a factor can be represented by

$$
Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1} \right) = \upsilon(x, \phi)\sigma_f^2. \quad\quad\quad\text{(C15)}
$$

In the remaining part of the proof, we focus on the relation between the amount of information about a sector factor, f_i, and the associated return variance. The same result applies to all other factors.

We first prove part A of Proposition 2. When the investor uses perfect Bayesian updating ($\phi = 1$), we can directly verify that $\upsilon_p(x, 1) + \upsilon_d(x, 1) = 1$. $\upsilon_p(x, 1)$ is the fraction of

uncertainty that is resolved early through information processing, and it increases with x. $v_d(x, 1)$ is the remaining unresolved fraction, and it decreases with x. Thus, if the investor processes more information, much uncertainty is resolved earlier, leaving less to be resolved later through dividends.

The total return variance generated by the factor

$$Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1}\right) = \left[\frac{1}{(1+r)^2}v_p(x, 1) + v_d(x, 1)\right]\sigma_f^2 \qquad (C16)$$

is determined by a weighted sum of the earlier resolution of uncertainty through information processing and the later resolution through the realization of dividends. We note that a discount factor, $\frac{1}{(1+r)^2}$, has been multiplied to the earlier resolution of uncertainty $v_p(x, 1)$, reflecting that early resolution of future uncertainty creates less return variation. Once we substitute out $v_d(x, 1)$ by $1 - v_p(x, 1)$, we obtain

$$Var\left(\frac{1}{1+r}\hat{f}_{i,t} + f_{i,t} - \hat{f}_{i,t-1}\right) = \left[1 - \frac{r(2+r)}{(1+r)^2}v_p(x, 1)\right]\sigma_f^2. \qquad (C17)$$

This expression equals to σ_f^2 if $r = 0$, and decreases with the amount of information, x, if $r > 0$. This result is summarized in Part A of Proposition 2.

We prove part B of Proposition 2 as follows. By the definition of v in equation (C14),

$$v(x, \phi) = \frac{1}{(1+r)^2}\frac{(1 - e^{-\phi x})^2}{1 - e^{-x}} + e^{-2\phi x} + (1 - e^{-\phi x})^2\frac{e^{-x}}{1 - e^{-x}}. \qquad (C18)$$

We analyze the properties of $v(x, \phi)$ around $x = 0$. Direct differentiation provides

$$\left.\frac{\partial v(x, \phi)}{\partial x}\right|_{x=0} = \frac{\phi}{(1+r)^2}[(\phi - 2)(1+r)^2 + \phi], \qquad (C19)$$

$$\left.\frac{\partial^2 v(x, \phi)}{\partial x^2}\right|_{x=0} = -\frac{\phi^2}{(1+r)^2}[2(\phi - 2)(1+r)^2 + (1+r)^2 + 2\phi - 1]. \qquad (C20)$$

In the derivation of these derivatives, we used the L'Hopital's rule.

If $\phi > 1 + \frac{(1+r)^2 - 1}{(1+r)^2 + 1}$, we can verify that

$$\left.\frac{\partial v(x, \phi)}{\partial x}\right|_{x=0} > 0, \quad \text{and} \quad \left.\frac{\partial^2 v(x, \phi)}{\partial x^2}\right|_{x=0} < 0. \qquad (C21)$$

Since v is a smooth function, there must exist a critical level, $\bar{x} > 0$, such that $v(x)$ is strictly increasing and concave for $\forall x \in [0, \bar{x})$. We determine $\bar{x}$ as the root to the equation

$$\partial v(x, \phi)/\partial x = 0. \tag{C22}$$

An analytical solution to this equation is not available. We numerically demonstrate the values of $\bar{x}$ in Appendix Appendix D for a wide range of model parameters.

Direct differentiation also provides that

$$\frac{\partial v(x, \phi)}{\partial \phi} = \frac{2x}{(1+r)^2(1-e^{-x})} \left[e^{-\phi x} - e^{-2\phi x} + (1+r)^2 \left(e^{-(1+\phi)x} - e^{-2\phi x} \right) \right] > 0. \tag{C23}$$

Thus, v increases with ϕ for any $x \geq 0$.

C.4 Proof to Proposition 3

We denote $x_h = \lambda_h \theta_h \kappa$ as the amount of market-level information processed by the investor, $x_{f,i} = \lambda_{f,i} \theta_i \kappa$ as the amount of information processed for sector i, and $x_{g,i,j} = \lambda_{g,i,j} \theta_i \kappa$ as the amount of information processed about firm j in sector i.

From equation (49), the return correlation between two firms from different sectors is

$$\rho = \frac{Var\left(\frac{1}{1+r}\hat{h}_t + h_t - \hat{h}_{t-1}\right)}{Var\left(\frac{1}{1+r}\hat{h}_t + h_t - \hat{h}_{t-1}\right) + Var\left(\frac{\hat{f}_{i,t}}{1+r} + f_{i,t} - \hat{f}_{i,t-1}\right) + Var\left(\frac{\hat{g}_{i,j,t}}{1+r} + g_{i,j,t} - \hat{g}_{i,j,t-1}\right)}$$

$$= \frac{v(x_h, \phi)\sigma_h^2}{v(x_h, \phi)\sigma_h^2 + v(x_{f,i}, \phi)\sigma_f^2 + v(x_{g,i,j}, \phi)\sigma_g^2}, \tag{C24}$$

where the function v is defined in equation (C14). From the factor structure in equation (4), the fundamental correlation between them is

$$\rho^D = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_f^2 + \sigma_g^2}. \tag{C25}$$

By rearranging the terms in equation (C24), we obtain

$$\frac{\rho}{1-\rho} = \frac{v(x_h, \phi)\sigma_h^2}{v(x_{f,i}, \phi)\sigma_f^2 + v(x_{g,i,j}, \phi)\sigma_g^2} \tag{C26}$$

According to Proposition 1, since $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$, we have that $x_h > x_{f,i} > x_{g,i,j}$. Furthermore, under the conditions of Proposition 2B, $v(x, \phi)$ is an increasing function of x.

39

Therefore, $\upsilon(x_h, \phi) > \upsilon(x_{f,i}, \phi)$ and $\upsilon(x_h, \phi) > \upsilon(x_{g,i,j}, \phi)$. Thus,

$$\frac{\rho}{1 - \rho} > \frac{\sigma_h^2}{\sigma_f^2 + \sigma_g^2} = \frac{\rho^D}{1 - \rho^D}, \tag{C27}$$

and $\rho > \rho^D$.

Following equation (51), the return correlation between any two firms in sector i is

$$\rho_i = \frac{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,i}, \phi)\sigma_f^2}{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,i}, \phi)\sigma_f^2 + \upsilon(x_{g,i,j}, \phi)\sigma_g^2}. \tag{C28}$$

The fundamental correlation in the dividends of these two firms is

$$\rho_i^D = \frac{\sigma_h^2 + \sigma_f^2}{\sigma_h^2 + \sigma_f^2 + \sigma_g^2}. \tag{C29}$$

Following a procedure similar to the one above, we can prove that $\rho_i > \rho_i^D$.

C.5   Proof to Proposition 4

We denote $x_{f,1} = \lambda_{f,1}\theta_1\kappa$ as the amount of information processed by the investor about the sector factor of sector 1, and $x_{f,1} = \lambda_{f,1}\theta_1\kappa$ as the amount of information processed about the sector factor of sector 2. The first-order condition in equation (42) for the investor's attention allocation to sector factors implies that if all factors receive positive attention allocation, then

$$\ln \mu_\kappa = \ln(\theta_1) + \ln(\phi\kappa n^2\sigma_f^2) - \phi x_{f,1}, \tag{C30}$$

$$\ln \mu_\kappa = \ln(\theta_2) + \ln(\phi\kappa n^2\sigma_f^2) - \phi x_{f,2}. \tag{C31}$$

By taking the difference between the two equations above, we obtain that

$$x_{f,1} - x_{f,2} = \frac{1}{\phi}(\ln\theta_1 - \ln\theta_2) > 0. \tag{C32}$$

We denote this difference by dx.

Since all the firm-specific factors in each sector have the same prior variance, they share the same amount of attention allocation. We denote $x_{g,1,j} = \lambda_{g,1,j}\theta_1\kappa$ as the amount of information processed by the investor about a firm-specific factor in sector 1, and $x_{g,2,j} = \lambda_{g,2,j}\theta_2\kappa$ as the amount of information processed by the investor about a firm-specific factor

in sector 2. The first-order condition in equation (43) for the investor's attention allocation to firm-specific factors implies that if all factors receive positive attention allocation, then

$$\ln \mu_\kappa = \ln(\theta_1) + \ln(\phi \kappa \sigma_g^2) - \phi x_{g,1,j}, \tag{C33}$$

$$\ln \mu_\kappa = \ln(\theta_2) + \ln(\phi \kappa \sigma_g^2) - \phi x_{g,2,j}. \tag{C34}$$

By taking the difference between the two equations above, we show that

$$x_{g,1,j} - x_{g,2,j} = \frac{1}{\phi}(\ln \theta_1 - \ln \theta_2) = x_{f,1} - x_{f,2} = dx > 0. \tag{C35}$$

We note that relative to sector 2, the amount of information that the investor processes for the sector factor and any firm-specific factor in sector 1 increases by the same amount, dx. Since the amount of firm-specific information is smaller ($x_{f,2} > x_{g,2,j}$ from Proposition 1), its increase is relatively larger.

To prove that return correlation of firms in sector 1 is lower ($\rho_1 < \rho_2$), we need only establish that

$$\frac{\rho_1}{1 - \rho_1} < \frac{\rho_2}{1 - \rho_2}, \tag{C36}$$

which, according to equation (C28), is equivalent to

$$\frac{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,1}, \phi)\sigma_f^2}{\upsilon(x_{g,1,j}, \phi)\sigma_g^2} < \frac{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,2}, \phi)\sigma_f^2}{\upsilon(x_{g,2,j}, \phi)\sigma_g^2}. \tag{C37}$$

By rearranging the terms, we can rewrite the inequality above as

$$\frac{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,1}, \phi)\sigma_f^2}{\upsilon(x_h, \phi)\sigma_h^2 + \upsilon(x_{f,2}, \phi)\sigma_f^2} - 1 < \frac{\upsilon(x_{g,1,j}, \phi)\sigma_g^2}{\upsilon(x_{g,2,j}, \phi)\sigma_g^2} - 1, \tag{C38}$$

which is further equivalent to

$$\frac{\upsilon(x_{f,2} + dx, \phi) - \upsilon(x_{f,2}, \phi)}{\upsilon(x_h, \phi)\sigma_h^2/\sigma_f^2 + \upsilon(x_{f,2}, \phi)} < \frac{\upsilon(x_{g,2,j} + dx, \phi) - \upsilon(x_{g,2,j}, \phi)}{\upsilon(x_{g,2,j}, \phi)}. \tag{C39}$$

Since $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$, Proposition 1 implies that $x_{g,2,j} < x_{f,2}$. Under the conditions of Proposition 2B, υ is an increasing and concave function of x for $x \in [0, \bar{x})$. Thus, we have

$$\upsilon(x_{g,2,j}, \phi) < \upsilon(x_{f,2}, \phi), \tag{C40}$$

$$\upsilon(x_{g,2,j} + dx, \phi) - \upsilon(x_{g,2,j}, \phi) > \upsilon(x_{f,2} + dx, \phi) - \upsilon(x_{f,2}, \phi) > 0. \tag{C41}$$

41

These two inequalities directly imply that inequality (C39) holds true. Thus, the return correlation of firms in sector 1 is lower. A key reason for this result is that in the more efficient sector, the investor not only processes more information for both sector- and firm-specific factors, but she also processes relatively more firm-specific information.

Next, we establish that stock prices in sector 1 are more informative. Using Lemma 3 in *C.3*, we can rewrite the inverse measures of stock price informativeness of firms in sectors 1 and 2, defined in equation (52), as

$$v_d(x_h, \phi)\sigma_h^2 + v_d(x_{f,1}, \phi)\sigma_f^2 + v_d(x_{g,1,j}, \phi)\sigma_g^2 \tag{C42}$$

and

$$v_d(x_h, \phi)\sigma_h^2 + v_d(x_{f,2}, \phi)\sigma_f^2 + v_d(x_{g,2,j}, \phi)\sigma_g^2, \tag{C43}$$

respectively. Lemma 3 shows that $v_d(x, \phi)$ is a decreasing function of x. Since $x_{f,1} > x_{f,2}$ and $x_{g,1,j} > x_{g,2,j}$ as in equation (C35), $v_d(x_{f,1}, \phi) < v_d(x_{f,2}, \phi)$ and $v_d(x_{g,1,j}, \phi) < v_d(x_{g,2,j}, \phi)$. Thus, the term in (C42) is smaller than that in (C43) and the stock prices of firms in sector 1 are more informative than those in sector 2.

C.6 Proof to Proposition 5

In period t, we denote the attention allocation to the market factor by $\lambda_{h,t}$, and the amount of information processed by $x_{h,t} = \lambda_{h,t}\theta_t\kappa$. Due to the symmetry assumption, all the sector factors receive an identical attention allocation, which we denote by $\lambda_{f,t}$; likewise, all firm-specific factors receive an identical attention allocation, denoted by $\lambda_{g,t}$. We specify $x_{f,t} = \lambda_{f,t}\theta_t\kappa$ as the amount of information processed for a sector factor and $x_{g,t} = \lambda_{g,t}\theta_t\kappa$ as the amount of information processed for a firm-specific factor. Furthermore, this symmetry implies that the return correlation between any two firms from different sectors is identical. We denote the return correlation as ρ_t.

To capture an increasing trend in θ, we consider the case that θ_t rises on even dates and stays constant on odd dates. Thus, $\cdots = \theta_{2t-1} < \theta_{2t} = \theta_{2t+1} < \theta_{2t+2} = \cdots$. On odd dates the investor's attention allocations stay the same from the previous date, thus simplifying our exposition. To prove the proposition, we need to show that the return correlation between

any two firms from different sectors decreases over time:

$$\rho_{2t+1} < \rho_{2t-1}. \tag{C44}$$

The first-order condition in equation (41) for the investor's attention allocation to the market factor implies that if all factors receive positive attention allocation, then

$$\ln \mu_{\kappa,2t+1} = \ln(\theta_{2t+1}) + \ln(\phi \kappa \sigma_h^2) - \phi x_{h,2t+1}, \tag{C45}$$

$$\ln \mu_{\kappa,2t-1} = \ln(\theta_{2t-1}) + \ln(\phi \kappa \sigma_h^2) - \phi x_{h,2t-1}. \tag{C46}$$

By taking the difference of these two equations, we obtain that

$$x_{h,2t+1} - x_{h,2t-1} = \frac{1}{\phi}[\ln(\theta_{2t+1}) - \ln(\theta_{2t-1}) - (\ln \mu_{\kappa,2t+1} - \ln \mu_{\kappa,2t-1})]. \tag{C47}$$

We denote the difference by dx.

Following a similar procedure, by using the first-order conditions in equations (42) and (43), we verify that if all factors receive positive attention allocation, then

$$x_{f,2t+1} - x_{f,2t-1} = x_{g,2t+1} - x_{g,2t-1} = x_{h,2t+1} - x_{h,2t-1} = dx. \tag{C48}$$

This result suggests that the changes in the amount of information processed for all factors are identical. Since all the attention allocations each period should sum up to one:

$$\lambda_{h,2t+1} + m\lambda_{f,2+1} + mn\lambda_{g,2t+1} = 1, \tag{C49}$$

$$\lambda_{h,2t-1} + m\lambda_{f,2t-1} + mn\lambda_{g,2t-1} = 1. \tag{C50}$$

These equations further imply that

$$x_{h,2t+1} + mx_{f,2+1} + mnx_{g,2t+1} = \kappa\theta_{2t+1}, \tag{C51}$$

$$x_{h,2t-1} + mx_{f,2t-1} + mnx_{g,2t-1} = \kappa\theta_{2t-1}. \tag{C52}$$

By taking the difference of these two equations above and substituting in dx from equation (C48), we can solve for dx:

$$dx = \frac{\kappa}{1+m+mn}(\theta_{2t+1} - \theta_{2t-1}) > 0. \tag{C53}$$

Thus, the changes in the amount of information processed for all factors over time are positive and identical. Since the amount of information processed for firm-specific factors is smaller than that of common factors, its percentage increase is larger.

To verify inequality (C44), we need only establish that

$$\frac{\rho_{2t+1}}{1 - \rho_{2t+1}} < \frac{\rho_{2t-1}}{1 - \rho_{2t-1}}, \tag{C54}$$

which, according to equation (C26), is equivalent to

$$\frac{\upsilon(x_{h,2t+1}, \phi)\sigma_h^2}{\upsilon(x_{f,2t+1}, \phi)\sigma_f^2 + \upsilon(x_{g,2t+1}, \phi)\sigma_g^2} < \frac{\upsilon(x_{h,2t-1}, \phi)\sigma_h^2}{\upsilon(x_{f,2t-1}, \phi)\sigma_f^2 + \upsilon(x_{g,2t-1}, \phi)\sigma_g^2}. \tag{C55}$$

By rearranging terms, we can rewrite the inequality above as

$$\frac{\upsilon(x_{h,2t+1}, \phi)\sigma_h^2}{\upsilon(x_{h,2t-1}, \phi)\sigma_h^2} - 1 < \frac{\upsilon(x_{f,2t+1}, \phi)\sigma_f^2 + \upsilon(x_{g,2t+1}, \phi)\sigma_g^2}{\upsilon(x_{f,2t-1}, \phi)\sigma_f^2 + \upsilon(x_{g,2t-1}, \phi)\sigma_g^2} - 1, \tag{C56}$$

which is equivalent to

$$\frac{\upsilon(x_{h,2t-1} + dx, \phi) - \upsilon(x_{h,2t-1}, \phi)}{\upsilon(x_{h,2t-1}, \phi)}$$
$$< \frac{[\upsilon(x_{f,2t-1} + dx, \phi) - \upsilon(x_{f,2t-1}, \phi)]\sigma_f^2/\sigma_g^2 + \upsilon(x_{g,2t-1} + dx, \phi) - \upsilon(x_{g,2t-1}, \phi)}{\upsilon(x_{f,2t-1}, \phi)\sigma_f^2/\sigma_g^2 + \upsilon(x_{g,2t-1}, \phi)}. \tag{C57}$$

Since $m^2\sigma_h^2 > \sigma_f^2$ and $n^2\sigma_f^2 > \sigma_g^2$, Proposition 1 implies that $x_{h,2t-1} > x_{f,2t-1} > x_{g,2t-1}$. Furthermore, under the conditions of Proposition 2B, υ is an increasing and concave function of x for $x \in [0, \bar{x})$. Thus, we have

$$\upsilon(x_{h,2t-1}, \phi) > \upsilon(x_{f,2t-1}, \phi) > \upsilon(x_{g,2t-1}, \phi) \tag{C58}$$

$$0 < \upsilon(x_{h,2t-1} + dx, \phi) - \upsilon(x_{h,2t-1}, \phi) < \upsilon(x_{f,2t-1} + dx, \phi) - \upsilon(x_{f,2t-1}, \phi), \tag{C59}$$

$$0 < \upsilon(x_{f,2t-1} + dx, \phi) - \upsilon(x_{f,2t-1}, \phi) < \upsilon(x_{g,2t-1} + dx, \phi) - \upsilon(x_{g,2t-1}, \phi). \tag{C60}$$

Inequalities (C59) and (C60) suggest that over time, return variances related to the investor's information processing for all factors are increased, but the magnitudes of these increases are different. In particular, the increase in the component related to firm-specific factors is largest, followed by the increase related to sector factors. The increase related to the market factor is smallest. This order determines the decrease in return correlation over time.

Inequality (C60) implies that the right-hand side of inequality (C57) is larger than

$$\frac{[\upsilon(x_{f,2t-1}+dx,\phi)-\upsilon(x_{f,2t-1},\phi)]\sigma_f^2/\sigma_g^2 + \upsilon(x_{f,2t-1}+dx,\phi)-\upsilon(x_{f,2t-1},\phi)}{\upsilon(x_{f,2t-1},\phi)\sigma_f^2/\sigma_g^2 + \upsilon(x_{g,2t-1},\phi)}$$

$$= [\upsilon(x_{f,2t-1}+dx,\phi)-\upsilon(x_{f,2t-1},\phi)]\frac{\sigma_f^2/\sigma_g^2 + 1}{\upsilon(x_{f,2t-1},\phi)\sigma_f^2/\sigma_g^2 + \upsilon(x_{g,2t-1},\phi)}$$

$$> \frac{\upsilon(x_{f,2t-1}+dx,\phi)-\upsilon(x_{f,2t-1},\phi)}{\upsilon(x_{f,2t-1},\phi)} \tag{C61}$$

where the last inequality is derived from inequality (C58) that $\upsilon(x_{f,2t-1},\phi) > \upsilon(x_{g,2t-1},\phi)$. Further applying inequalities (C59) and (C58) for the numerator and denominator of (C61), respectively, we can verify that the right-hand-side term in (C61) is larger than the left-hand side of inequality (C57). Thus, inequality (C57) holds, as well as inequality (C44).

Using a similar method, we can also prove that the return correlation between any two firms in the same sector also decreases over time as the parameter θ increases.

C.7   Proof to Proposition 6

By substituting the linear price function in Theorem 1 and the factor structure in equation (4) into an asset's excess return in equation (44), we can divide the first-order autocovariance of the asset's return into three components corresponding to the three fundamental factors involved:

$$Cov(q_{i,j,t}, q_{i,j,t+1}) = Cov\left(\frac{\hat{h}_t}{1+r} + h_t - \hat{h}_{t-1}, \; \frac{\hat{h}_{t+1}}{1+r} + h_{t+1} - \hat{h}_t\right)$$

$$+ Cov\left(\frac{\hat{f}_{i,t}}{1+r} + f_{i,t} - \hat{f}_{i,t-1}, \; \frac{\hat{f}_{i,t+1}}{1+r} + f_{i,t+1} - \hat{f}_{i,t}\right)$$

$$+ Cov\left(\frac{\hat{g}_{i,j,t}}{1+r} + g_{i,j,t} - \hat{g}_{i,j,t-1}, \; \frac{\hat{g}_{i,j,t+1}}{1+r} + g_{i,j,t+1} - \hat{g}_{i,j,t}\right) \tag{C62}$$

Here, we analyze the component related to the firm-specific factor as an illustration. We can derive the other two components using the same method.

From the independence assumption of the fundamental factors across periods and the associated Bayesian learning processes, $g_{i,j,t}-\hat{g}_{i,j,t-1}$ is independent of $\hat{g}_{i,j,t+1}$ and $g_{i,j,t+1}-\hat{g}_{i,j,t}$, and $\hat{g}_{i,j,t}$ is independent of $\hat{g}_{i,j,t+1}$. Thus, we obtain

$$Cov\left(\frac{\hat{g}_{i,j,t}}{1+r} + g_{i,j,t} - \hat{g}_{i,j,t-1}, \; \frac{\hat{g}_{i,j,t+1}}{1+r} + g_{i,j,t+1} - \hat{g}_{i,j,t}\right)$$

$$
\begin{aligned}
&= \; Cov\left(\frac{\hat{g}_{i,j,t}}{1+r}, \; g_{i,j,t+1} - \hat{g}_{i,j,t}\right) \\
&= \; \frac{1}{1+r}Cov\left(\hat{g}_{i,j,t}, \; g_{i,j,t+1} - \hat{g}_{i,j,t}\right) \\
&= \; \frac{1}{2(1+r)}\left[Var(\hat{g}_{i,j,t} + g_{i,j,t+1} - \hat{g}_{i,j,t}) - Var(\hat{g}_{i,j,t}) - Var(g_{i,j,t+1} - \hat{g}_{i,j,t})\right] \quad \text{(C63)} \\
&= \; \frac{1}{2(1+r)}\left\{Var(g_{i,j,t+1}) - \left[Var(\hat{g}_{i,j,t}) + Var(g_{i,j,t} - \hat{g}_{i,j,t-1})\right]\right\}. \quad \text{(C64)}
\end{aligned}
$$

In deriving equation (C63), we use the following formula:

$$
Cov(x,y) = \frac{1}{2}[Var(x+y) - Var(x) - Var(y)]. \quad \text{(C65)}
$$

We note that $Var(g_{i,j,t+1}) = \sigma_g^2$. Based on Appendix $C.3$, $Var(\hat{g}_{i,j,t}) + Var(g_{i,j,t} - \hat{g}_{i,j,t-1})$ is the return variance related to the firm-specific factor when the discount rate is zero. If we rewrite υ defined in equation (C14) as a function $\upsilon(x, \phi, r)$ where x is the amount of information processed about a factor, ϕ is the investor's overconfidence parameter, and r is the discount rate, we have

$$
Var(\hat{g}_{i,j,t}) + Var(g_{i,j,t} + \hat{g}_{i,j,t-1}) = \sigma_g^2 \, \upsilon(x_{g,i,j}, \phi, 0), \quad \text{(C66)}
$$

where $x_{g,i,j}$ is the amount of information that the investor processes about the firm-specific factor. Thus,

$$
Cov\left(\frac{\hat{g}_{i,j,t}}{1+r} + g_{i,j,t} - \hat{g}_{i,j,t-1}, \; \frac{\hat{g}_{i,j,t+1}}{1+r} + g_{i,j,t+1} - \hat{g}_{i,j,t}\right) = \frac{\sigma_g^2}{2(1+r)}[1 - \upsilon(x_{g,i,j}, \phi, 0)]. \quad \text{(C67)}
$$

The source of the autocovariance is the investor's overreaction to her information. Using equation (C18), we have

$$
\upsilon(x, \phi, 0) = 1 + \frac{2(1 - e^{-\phi x})}{1 - e^{-x}}(e^x - e^{-\phi x}) \geq 1, \quad \text{(C68)}
$$

and that the equality holds only when $x = 0$. Thus, if the investor does not attend to any information related to the factor, there is no overreaction and therefore no autocovariance in the return related to the factor, either. According to equation (43), this situation occurs when $\theta_i \leq \mu_\kappa/(\phi\kappa\sigma_g^2)$.

Furthermore, under the conditions of Proposition 2B, $\upsilon(x, \phi, r)$ increases with x. Thus, equation (C67) implies that in this region, the negative autocovariance caused by the investor's overreaction decreases with the amount of firm-specific information processed.

C.8 Proof to Proposition 7

The information value of the ignored signal, $\nu_{i,j,t}$, depends on its ability to resolve the remaining uncertainty in prices. Thus, we can measure it by its correlation with $g_{i,j,t+1} - \hat{g}_{i,j,t}$, the remaining uncertainty in the firm's fundamental factor. We denote $x = \lambda_{g,i,j,t}\theta_i\kappa$ as the amount of firm-specific information processed by the investor. By substituting in $\hat{g}_{i,j,t}$ from equation (29), we obtain

$$g_{i,j,t+1} - \hat{g}_{i,j,t} = e^{-\phi x}(g_{i,j,t+1} - \bar{g}) - (1 - e^{-\phi x})\epsilon_{g,i,j,t}, \tag{C69}$$

where $\epsilon_{g,i,j,t}$ is the noise in the information processed by the investor in equation (12), and it is independent of $\epsilon_{\nu,i,j,t}$, the noise in the ignored signal. Then, we can compute the correlation between $\nu_{i,j,t}$ in equation (53) and $g_{i,j,t+1} - \hat{g}_{i,j,t}$ as

$$Corr = \frac{e^{-\phi x}\sigma_g^2}{\sqrt{[e^{-2\phi x}\sigma_g^2 + (1 - e^{-\phi x})^2\eta_g^2](\sigma_g^2 + \sigma_\nu^2)}} \tag{C70}$$

where $\eta_g^2 = \sigma_g^2/(e^x - 1)$ is the variance of $\epsilon_{g,i,j,t}$. We rearrange the correlation as

$$Corr = \frac{\sigma_g}{\sqrt{[1 + \frac{(e^{\phi x}-1)^2}{e^x-1}](\sigma_g^2 + \sigma_\nu^2)}}. \tag{C71}$$

Direct differentiation shows that $\frac{(e^{\phi x}-1)^2}{e^x-1}$ is a strictly increasing function of x. Thus, $Corr$ decreases with the amount of firm-specific information processed by the investor.

Appendix D   Numerical illustration of relation between return variance and information

In Section *5.1*, we discuss the relation between return variance and the amount of information processed by the investor about a certain fundamental factor. Proposition 2B shows that when the investor is sufficiently overconfident, the return variance increases with the amount of processed information when it is below a threshold. In this appendix, we illustrate this relation and the threshold using several numerical examples.

According to equation (C15), the return variance related to a factor is determined by the factor's fundamental variance multiplied by a function $v(x, \phi)$, which captures the return

47

variance amplification caused by the investor's overreaction to her information. $x = \lambda_{f,i,t}\theta_i\kappa$ is the amount of information processed about the factor and ϕ is the investor's overconfidence parameter.
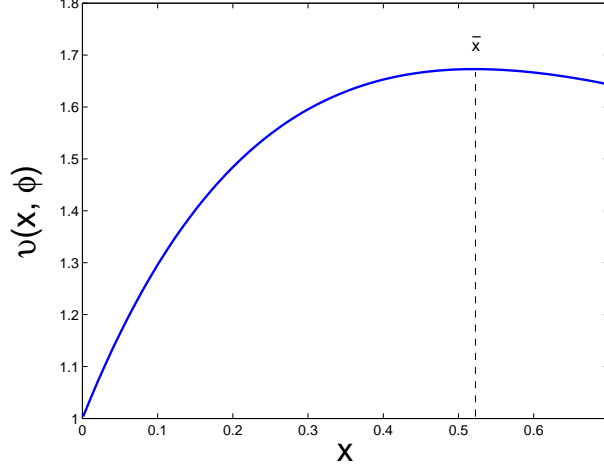

Figure 1: The relation between return variance amplification, $v(x, \phi)$, and the amount of processed information, x. The figure uses the following parameters: $\phi = 2$ and $r = 5\%$.

Figure 1 demonstrates a typical shape of the v function based on $\phi = 2$ and a discount rate $r = 5\%$. v increases with the amount of information when it is below a threshold $\bar{x}$, which has a value of 0.521 in this example, and then becomes decreasing with the amount of information when it passes above $\bar{x}$. For the main part of our analysis, we restrict the amount of information processed by the investor to be below the threshold level. Based on equations (15) and (16), the economic interpretation of $\bar{x}$ corresponds to a threshold level in the logarithmic reduction in the variance of the investor's belief about the underlying fundamental factor:

$$\bar{x} = \log(\sigma^2_{prior}/\sigma^2_{posterior}). \tag{D1}$$

A value of 0.521 represents a percentage reduction in the variance of investor belief by

$$\Delta\sigma^2/\sigma^2 \equiv \frac{\sigma^2_{posterior} - \sigma^2_{prior}}{\sigma^2_{prior}} = e^{-\bar{x}} - 1 = -40.6\%, \tag{D2}$$

which is a substantial amount to be achieved by individual investors' learning processes.

48

Table 1 provides the values of the threshold ($\bar{x}$) and the corresponding percentage reduction in the variance of investor belief ($\Delta\sigma^2/\sigma^2$) by varying the values of the investor's overconfidence parameter (ϕ) and the discount rate (r). Panel A shares a discount rate $r = 0$, and Panel B shares a discount rate $r = 5\%$. For the investor's overconfidence parameter, we choose a wide range of values from 1.2 to 2.4. Doing so implies that the investor exaggerates her information processing ability by these multiples. Across all these values, the threshold ($\bar{x}$) stays between 0.449 and 0.857, and the corresponding reduction in belief variance is between 36% and 58%. All these values are high for individual investors' learning processes.

Also, we note that the lower bound for the investor's overconfidence parameter ϕ in Proposition 2B, inequality (47), is rather modest. For example, when the discount rate $r = 5\%$, the lower bound of ϕ corresponds to a value of 1.049.

Table 1: The threshold in amount of information

This table provides the values of the threshold ($\bar{x}$) in the amount of information and the corresponding percentage reduction in the variance of investor belief ($\Delta\sigma^2/\sigma^2$) for different values of the investor's overconfidence parameter (ϕ) and the discount rate (r).

	Panel A: $r = 0$			Panel B: $r = 5\%$	
ϕ	$\bar{x}$	$\Delta\sigma^2/\sigma^2$	ϕ	$\bar{x}$	$\Delta\sigma^2/\sigma^2$
1.2	0.857	-57.6%	1.2	0.646	-47.6%
1.4	0.751	-52.8%	1.4	0.657	-48.2%
1.6	0.669	-48.8%	1.6	0.613	-45.8%
1.8	0.603	-45.3%	1.8	0.565	-43.2%
2.0	0.549	-42.2%	2.0	0.521	-40.6%
2.2	0.504	-39.6%	2.2	0.483	-38.3%
2.4	0.466	-37.2%	2.4	0.449	- 36.2%

References

Barber, B., Odean, T., 2003. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Unpublished working paper. UC Davis and UC Berkeley.

Barber, B., Odean, T., Zhu, N., 2003. Systematic noise. Unpublished working paper. UC Davis and UC Berkeley.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 48, 307-343.

Barberis, N., Shleifer, A., 2003. Style investing. Journal of Financial Economics 68, 161-199.

Barberis, N., Shleifer, A., Wurgler J., 2005. Comovement. Journal of Financial Economics 75, 283-317.

Barberis, N., Thaler, R., 2003. A survey of behavioral finance. In: Constantinides G., Harris, M., Stulz, R. (Ed.), Handbook of the Economics of Finance, North-Holland, Amsterdam.

Bernardo, A., Welch, I., 2001. On the evolution of overconfidence and entrepreneurs. Journal of Economics and Management Strategy, 10, 301-330.

Boyer, B., 2004. Style investing and comovement among stocks with similar book-to-market ratios. Unpublished working paper. Brigham Young University.

Campbell, J., Lettau, M., Malkiel, B., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56, 1-43.

Cooper, M., Dimitrov O., Rau R., 2001. A rose.com by any name. Journal of Finance 56, 2371-2388.

Cover, T., Thomas, J., 1991. Elements of Information Theory. John Wiley & Sons, Inc., New York.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839-1885.

Della Vigna, S., Pollett, J., 2003. Attention, demographics, and the stock market. Unpublished working paper. UC Berkeley and Harvard University.

Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does greater firm-specific return variation mean more or less informed stock pricing? Journal of Accounting Research 41, 797-836.

Durnev, A., Morck, R., Yeung R., 2004. Does firm-specific information in stock prices guide capital budgeting? Journal of Finance 59, 65-105.

Einhoen, H., 1980. Overconfidence in judgement. New Directions for Methodology of Social and Behavioral Science 4, 1-16.

French, K., Roll, R., 1986. Stock return variances: the arrival of information and the reaction of traders. Journal of Financial Economics 17, 5-26.

Gabaix, X., Laibson, D., 2003. Bounded rationality and directed cognition. Unpublished working paper. MIT and Harvard University.

Gabaix, X., Laibson, D., Moloche, G., Weinberg, S., 2003. The allocation of attention: theory and evidence. Unpublished working paper. MIT and Harvard University.

Gervais, S., Odean, T., 2001. Learning to be overconfident. Review of Financial Studies 14, 1-27.

Grossman, S., Stiglitz, J., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393-408.

Hirshleifer, D., 2001. Investor psychology and asset pricing. Journal of Finance 56, 1533-1597.

Hirshleifer, D., Teoh, S. H., 2003. Limited attention, financial reporting and disclosure. Journal of Accounting and Economics 36, 337-386.

Hirshleifer, D., Teoh, S. H. 2004. Limited investor attention and earnings-related under- and over-reactions. Unpublished working paper. Ohio State University.

Hirshleifer, D., Lim S., Teoh, S. H., 2003. Disclosure to a credulous audience: The role of limited attention. Unpublished working paper. Ohio State University.

Hirshleifer, D., Hou, K., Teoh, S. H., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets? Journal of Accounting and Economics 38, 297-331.

Hong, H., Stein, J., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143-2184.

Hong, H., Stein, J., 2003. Simple models and paradigm shifts. Unpublished working paper. Princeton University and Harvard University.

Hong, H., Torous, W., Valkanov, R., 2003. Do industries lead the stock market? Unpublished working paper. Princeton University and UCLA.

Hou, K., Moskowitz, T., 2003. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies, forthcoming.

Huberman, G., Regev, T., 2001. Contagious speculation and a cure for cancer: a non-event that made stock prices soar. Journal of Finance 56, 387-396.

Kahneman, D., 1973. Attention and Effort. Prentice Hall, New Jersey.

Kaniel, R., Saar, G., Titman, S., 2004. Individual investor sentiment and stock returns. Unpublished working paper. Duke University, New York University and University of Texas.

Kumar, A., Lee, C., 2003. Mass psychology and return comovement: the case of retail trades. Unpublished working paper. University of Notre Dame and Cornell University.

Kyle, A., Wang, A., 1997. Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? Journal of Finance 52, 2073-2090.

Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies 1, 41-66.

Lucas, R., 1978. Asset prices in an exchange economy. Econometrica 46, 1429-1445.

Merton, R., 1971. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3, 373-413.

Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: Why do emerging markets have synchronous stock Price movements? Journal of Financial Economics 58, 215-260.

Mullainathan, S., 2002. Thinking through categories. Unpublished working paper. MIT.

Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. Journal of Finance 53, 1887-1934.

Pashler, H., Johnston J., 1998. Attentional limitations in dual-task performance. In: Pashler, H., (Ed.), Attention, Psychology Press.

Peng, L., 2005. Learning with information capacity constraints. Journal of Financial and Quantitative Analysis 40, 307-330.

Poterba, J., Summers, L., 1988. Mean reversion in stock returns: evidence and implications. Journal of Financial Economics 22, 27-59.

Pindyck, R., Rotemberg, J., 1993. The comovement of stock prices. Quarterly Journal of Economics 108, 1073-1104.

Roll, R., 1988. R^2. Journal of Finance 43, 541-566.

Scheinkman, J., Xiong, W., 2003. Overconfidence and speculative bubbles. Journal of Political Economy 111, 1183-1219.

Shiller, R., 1989. Comovement in stock prices and comovement in dividends. Journal of Finance 44, 719-729.

Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance 52, 35-55.

Sims, C., 2003. Implications of rational inattention. Journal of Monetary Economics, 50, 665-690.

Sloan, R., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Accounting Review 71, 289-315.

West, K., 1988. Dividend innovations and stock price volatility. Econometrica 56, 37-61.

Wurgler, J., 2000. Financial markets and the allocation of capital. Journal of Financial Economics 58, 187-214.

Van Nieuwerburgh, S., Veldkamp, L., 2004. Information acquisition and portfolio under-diversification. Unpublished working paper. New York University.

Verrecchia, R., 1982. Information acquisition in a noisy rational expectations economy. Econometrica 50, 1415-1430.

Yantis, S., 1998. Control of visual attention. In: Pashler, H., (Ed.), Attention, Psychology Press.