A Rational Inattention Perspective on Equilibrium Asset Pricing under Heterogeneous Information with Structural Breaks and Market Efficiency

Steve, Heinke and Niels, Warmuth

University of Zurich, Munich University of Technology

2016

A Rational Inattention Perspective on Equilibrium Asset Pricing under Heterogeneous Information with Structural Breaks and Market Efficiency

By Steve Heinke and Niels J. Warmuth*

In this paper we present a new model of how information travels within financial markets and present empirical evidence that the concept of attention driven information efficiency is more conjugate with market data as compared to the prevailing concept of efficient markets. Augmenting our model by a shift component made it possible to explain shifts in asset prices by a lack of attention on small permanent changes in the fundamentals. This can also be seen as a micro-level explanation of the momentum effect. By a further augmentation of the model through the introduction of heterogeneous information processing capacities we are able to give a fundamental interpretation of the financial services industry as providers of information processing capacity. Moreover, the burst of the housing bubble in the US and the successful bet of John Paulson against it are shown to be prime empirical examples of our framework.

JEL: C60, D80, D83, D84, D92, G12, G14
Keywords: Limited attention, asset pricing, rational inattention, momentum trading

* Heinke, Steve: University of Zurich, Wintherturerstrasse 30 - 8057 Zurich, steve.heinke@econ.uzh.ch. Warmuth, Niels Jakob: Munich University of Technology, Arcisstrasse 21 - 80333 Munich, niels.warmuth@wi.tum.de. Acknowledgements: We are grateful to Michael Allman, Marios Angeletos, Andreas Hefti, Josef Falkinger, Igor Letina, Sabrina Studer, Thomas Schmid, Christoph Jäckel, Christoph Kaserer, and to seminar participants at the Munich University of Technology, University of Zurich, and the Zurich Workshop on Economics 2013 in Schaffhausen (Switzerland).

I. Introduction

In a world of perfect markets, asset prices should fully reflect all available information on the future cash flows associated with holding this asset (Fama, 1970). Consequently, the riskiness of the asset ought to be the only factor a rational agent takes into account when he makes his trading decision. Nowadays information is available in massive amounts (Hefti and Heinke, 2015), but still the existence of profitable trading strategies beside risk challenges the notions of market efficiency and asset pricing based solely on rational expectation.[1] Moreover, the existence of funds spending vast amount of money on information processing and data mining just to trade profitable on this informational advantage, underlines that the availability of information alone is not the bottleneck when it comes to expectation formation. The abundance of information leads to a “scarcity of attention” goes a long (Simon, 1955) and and the relevant question for an investor changes to how to deal with this overwhelming amount of information. In other words, in an information rich environment, attention to each piece of information becomes a scarce resource.[2]

Two interlinked questions arise from the previous arguments:: First, why does the efficient market hypothesis not always holds? Second, why do arbitrage opportunities exist? Theory based answers to these questions can be mainly divided into two strands of the literature. The first deals with the question what hinders rational arbitrageurs to take advantage of mispricing in the market and to equalize them. Thus these concepts are mainly focused on institutional barriers.[3] The second strand investigates the specific nature of why agents fail to act fully rational. Common approaches are to assume special preferences[4] or biased beliefs.[5] While these approaches explain some of the empirically found anomalies, they have the drawback that their set-up is somewhat ad-hoc and that empirical findings tend to find a diminishing effect of most anomalies in the long run, which is not implied by most of these models (Fama, 1998).

This paper gives an attention based answer to both questions, by combining the concept of rational inattention (Sims, 1998, 2003, 2005) with the well developed overlapping generation framework of asset pricing with heterogeneous informed agents (Biais et al., 2010). We add heterogeneity among agents with respect to signals and their information processing capacity constraints to a model of a disaggregated economy with agents with an infinite horizon..

[1] See e.g. Jegadeesh and Titman (1993), Lakonishok et al. (1994), Hong et al. (2000) and Shleifer (2000) for an overview on behavioral finance.

[2] See e.g. the models of Falkinger (2007, 2008); Hefti (2013) and for an overview Hefti and Heinke (2015).

[3] An extensive review on these limits to arbitrage arguments can be found in Brunnermeier (2009).

[4] See e.g. Barberis et al. (2001)

[5] See among others Barberis et al. (1998) and Daniel et al. (1998)

Our first main finding stems from introducing uncertainty about the mean of a dividend process. This allows us distinguish between long-term and short-term factors of the assets dividend process.We study the influences of this distinction on the equilibrium price and the attention allocation, as well as on how the information seep into the economy. This yields an explanation for the empirical finding that shifts in the long-term fundamentals need time till they are fully reflected in the price. This explains momentum trading strategies as an outcome of rational actions for the price and that asset prices fully reflect changes in fundamentals only in the long run. The second result of this paper follows from allowing for heterogeneity in the information processing constraints. In this framework we discuss how algorithmic and high-frequency traders turn their capability of processing enormous amounts of data into an informational advantage, which generates excess returns. From a more general perspective, a similar argument can be used to explain the existence of financial services, since pooling resources allows for specialization in information gathering and thus allows for a more efficient use of the scarce resource attention capacities. This efficiency gain can be expressed in terms of excess returns and thus justifies paying other people to let them do the investment decision. The third finding is, that our model in its basic form is supported by the data. In particular we discuss some trading strategies of hedgefunds before the burst of the US sub prime bubble as an empirical case study example of the more advanced form of our model. Moreover, we argue that thinking the attention approach consequently through, the concept of market efficiency and its testing by event-studies is contingent on the level of attention on these event.

This study is nested on the overlap of two strands of the theoretical literature. On the one hand we built on the concept of rational inattention in order to model information acquisition and attention allocation. On the other hand we use an overlapping generation framework which allows us to model information aggregation processes within a competitive asset market. This short **literature review** will give an overview on what rational inattention is about and portray the overlapping generation models framework.

When it comes to **modelling attention** one has to keep in mind that the attention of humans is driven by two fundamental psychological forces (Pashler, 1998). One possibility is that attention is directed on the affective level, also called bottom-up or stimulus driven attention. In a newspaper based example this would be the colour or font-size of the article, which draws the attention towards the information. Another possible force is the cognitive aspect, also known as goal-driven or top-down attention. In the newspaper example the color of font size does not matter towards which information is processed, the only factor that matters whether the subject is interested in this topic is the likely benefit to have the information contained in the article. Usually both level interfere with each other, and thus play a role in the process how the information enters the deliberation.

Depending on the context of the decision to be made each aspect can be weighted differently. Since asset pricing is a cognitive task and very different from impulsive shopping for example we think that the rational inattention framework is the right one in this case..

Models of rational inattention focus only on goal driven aspect of the attention allocation processes. Thus the agent has an active role and full control when it comes to decide which signal he receives (Hefti, 2013). Returning to the newspaper example consider a fund manager who is in charge of the investment decision. Every morning he gets a newspaper and can decide on how much time he is going to spend on reading the newspaper itself as well as on how much of this time he is going to devote on the economics, finance, and politics section of this newspaper. Suppose that the more of the newspaper he reads, the better will be his idea of what is going on in the world and therefore he is more likely to make good investment decision, but less spare time remains for doing other relevant tasks. Given there is less time available than required to read the entire newspaper, he faces the problem of how to allocate the given time over the different subsections of the newspaper. For an investment decision reading the finance section might be most relevant. Nevertheless, reading the politics or general economics section might also be important since certain topics such as general economic policies, decision on warfare, strikes and so on will be discussed there, which could potentially matter for the investment decision as well. Under the assumption that the investor knows the structure of his newspaper, he can judge the "average" information potential of each section in advance, and thus decides rationally on the allocation of a given time over the subsections as well as on the total reading time by weighting the expected benefits of the optimal reading strategy against the costs of doing something else. It is important to understand that the investor allocates his given reading time, meaning his mental resources, only according to the ex-ante expected information content of the subsections.[6]

Sims[7], was among the first to formalize such an allocation problem, using the concept of entropy[8] as the measure of informativeness of a channel. In most parts of the literature a channel is simply a signal that is correlated with the future state of the world. DeOliveira et al. (2013) generalizes this idea by a decision theoretic foundation. The whole problem itself can be seen as a two-stage decision making process. In stage II the agent solves a standard utility maximizing problem for any posterior distribution, meaning for any belief about the states of the world. The objective of stage I is to choose the optimal channel/signal that maximize the information value with respect to the subjective costs. In our example the investor decides on his optimal trading strategy given some allocation of time over the subsections of the newspaper in stage II. At stage I

[6] Hence the investor's attention allocation is invariant to fancy headlines, pictures or report framing, which would be the assumption of stimulus-driven attention models. See Hefti and Heinke (2015) for an overview on models with stimulus driven attention, with a similar illustrative newspaper example.

[7] Sims (1998, 2003, 2005)

[8] From Shannon's information theory

he chooses overall reading time and attention allocation such that his expected utility of choosing an investment plan is maximized, by weighting up the utility costs of additional reading time against the marginal benefit of obtaining better information, meaning a better posterior, on the state of the world and thus on the prospects of his investment decision. Therefore the agent is totally rational in the sense that he optimizes both over information acquisition and investment actions. Due to the fact that information is subjectively costly, he will be inattentive to information that is not ex-ante promising to be useful relative to information costs. Most of the modelling done in rational inattention focuses on the variation of the subjective cost function.[9] Most of the literature mentioned above as well as our model builds the subjective cost function on the information theoretic concept of mutual information.[10]

The rational inattention approach has been applied to diverse macroeconomic sub-fields such as sticky prices, see Sims (2003) and Woodford (2009), differences in the price reactions due to different shocks, see Mackowiak and Wiederholt (2009) and Matejka and McKay (2012), understanding the forward discount puzzle of the uncovered interest rate parity condition, see Bacchetta and VanWincoop (2005), business cycles, see Mackowiak and Wiederholt (2009), and consumption choice with asymmetric responses to wealth shocks, see Tutino (2013), or in finance for studying portfolio allocation decision, see Peng (2005), understanding home bias, see Mondria et al. (2010), sectoral instead of firm specific learning, see Peng and Xiong (2006), and under diversification, see VanNieuwerburgh and Veldkamp (2008). Other fields in economics where the rational inattention approach has been used are coordination games, see Hellwig and Veldkamp (2009), and business studies by investigating a team production problem with task specialization resulting in an emergency of organizational leadership, see Dessein et al. (2013).[11]

The **overlapping generation framework** is used within the field of asset pricing mostly when one wants to discuss the information aggregation process as Hellwig (1980) did in his seminal work in which he studied the implications for the information contained in the price when each agent has a different piece of information. He concluded that in large markets only the common element of

[9]See for example Hellwig et al. (2012) for a discussion on information choice technologies.

[10]The main idea behind mutual information is, that the agent wants to know more about the normally distributed random variable X with variance σ_X^2, but can only observe the signal s, where X and s have a multivariate normal distribution with conditional variance $\sigma_{X|s}^2$ of X. The unconditional entropy $H(\cdot)$ of X is given by $H(X) = \frac{1}{2}log_2(2\pi e\sigma_X^2)$. This can be interpreted as a measure of uncertainty. The conditional entropy of X after observing the signal s is $H(X|s) = \frac{1}{2}log_2(2\pi e\sigma_{X|s}^2)$. With these measures at hand one can calculate the mutual information the signal s contains about the random variable X and vice versa, by deducting the conditional entropy from the unconditional one $I(X;s) = H(X) - H(X|s)$. Equipped with the quantification of mutual information, limited attention capacities are modelled by a bound κ on its per period average: $\mathbb{I}(X;s) \leq \kappa$

[11]Veldkamp (2011) is a comprehensive source for further applications of the rational inattention concept.

information that is known to many agents is reflected in the equilibrium price. Apart from this the overlapping generations framework the agents interaction in a market is explicitly modeld and specifically for the financial market this model therefore represents a realistic approach. Biais et al. (2010) rely on these insights and study the equilibrium prices and portfolio selection when there are agents with asymmetric information sets in the market. Their setting the less informed agents face a winners curse problem and has to take this into considerations when deciding on the portfolio selection. Indexing fails and there are possibilities to outperform the market.

We build on the frameworks of Biais et al. (2010) and Sims (1998), since we are interested in heterogeneous informed agents and their information choices, how these choices affect asset prices, and how agents interact with each other. The resulting model incorporates heterogeneity of agents in two ways, at first in the information capacity constraint of each agent and second in the signal itself. We take advantage of both features of the model and are able to derive interesting results showing that attention is a relevant factor when one wants to understand anomalies in asset price dynamics within a competitive market.

The rest of this article is organised as follows. Section II introduces the simplest versions of our model and theoretical framework highlighting the basic mechanisms behind it. Subsequently, in section III, we extend the basic version of the model to provide testable implications with respect to market efficiency and to incorporate shifts in our framework. Section IV discusses the model with heterogeneous agents. Section V concludes.

II. Model

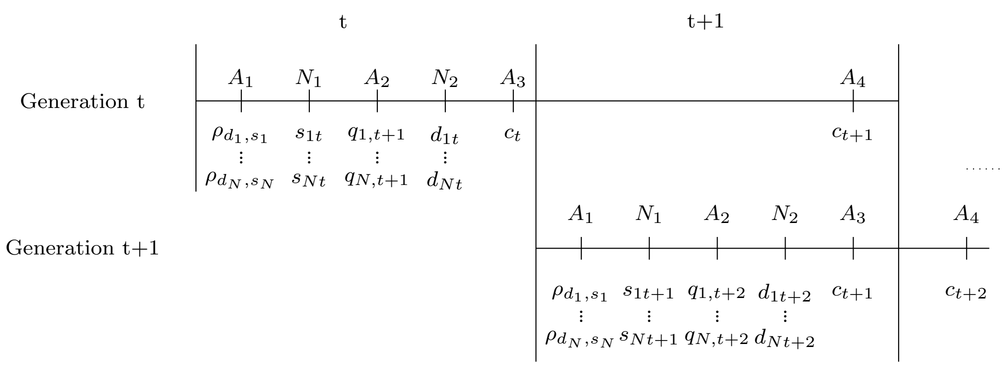
Consider an economy with N assets and the dividend flow of each asset $n \in N$ follows a stochastic process with a deterministic mean μ_n and variance σ_n^2. Thus the dividend stream can be described by,

$$d_{n,t} = \mu_n + \sigma_n \epsilon_{nt}, \tag{1}$$

where $\epsilon_{nt} \sim N(0,1)$ represents the normally distributed random component, with mean 0 and variance 1. In each period t the old generation $t-1$ is already in the economy owning the assets and a new generation t is born consisting of a continuum of identical agents distributed uniformly on the unit interval with a constant population mass of one. The agents of generation t can buy the asset and earn the dividend, which will be left from paying for the assets can be consumed. In the next period $t+1$ generation t resells the asset and consume the received cash. Thus owning an asset means saving for the second period consumption. Before buying the asset, a member of generation t can only receive noisy signals about the dividends by informing themselves. Since they are limited by their information processing capabilities they have to decide on how thoroughly they want inform themselves, meaning how noisy the signal for each asset should be.

This implies the following order of events (see figure 1). Each agent $i \in [0,1]$ of the young generation t first decides how to allocate his attention, then he receives his information on the dividends in period t and decides on his trading strategy q^i_{t+1}. After this trading takes place, while the old generation $t-1$ will sell all its assets to finance their consumption in period t, the young generation t will buy the assets in order so save for consumption in its second period. The prices clear markets. Finally, the dividend realizes and the residual of dividends less payments for the bought assets will be consumed. In period $t+1$ generation $t-1$ leaves the economy, generation t will be the old generation and a new generation $t+1$ is born, repeating the procedure above.

Figure 1. Timing of Events



Graphically representation of timing of events in this the overlapping generation framework. The life of generation t and t+1 are represented by the two time lines. $A_{1,2,3,4}$ mark events where the generation has to take an action or make a decision. $N_{1,2}$ stand for an occurrence by nature. s_{nt} stands for the signal belonging to asset n in time period t. ρ_{d_n,s_n} is the correlation between signal n and dividend of asset n. q_{nt+1} is the number of asset n hold from period t to period t+1. c_t and c_{t+1} are the consumptions associated with the investment in period t and t+1 for generation t.

The signal structure of asset n in period t agent i chooses to observe is taken from the set of all possible signal structures Γ and consists of the future dividend, equation (1), plus noise $\tilde{\sigma}^i_n \psi^i_{nt}$, where $\tilde{\sigma}^i_n$ is the scaling parameter of the noise and ψ^i_{nt} is normally distributed with mean 0 and variance 1. The signal precision is a function of the amount of information contained in the signal about the dividend, measured by the average mutual information the signal contains about the dividend. Since the agent i only has limited capacities to process information, there exists an upper bound κ^i on the amount of information processed $\mathbb{I}(\cdot)$. Thus κ^i can be thought of as the maximum information processing capacity. We will refer to (3) as information processing constraint.

$$s^i_{n,t} = \mu_n + \sigma_n \epsilon_{nt} + \tilde{\sigma}^i_n \psi^i_{nt} \quad (2)$$

$$\mathbb{I}\left(\{d_t\}; \{s^i_t\}\right) \leq \kappa^i \quad (3)$$

Where s_t^i is the vector of all signals chosen by agent i and d_t is the vector of the stochastic processes of all assets' dividends. The agent's inter-temporal rate of substitution is given by β.

With this notation at hand one can describe the two-stage decision problem of the agent:

$$\max_{s_t^i \in \Gamma} \mathbb{E}\left[u(c_t^i; s_t^i) + \beta u(c_{t+1}^i; s_t^i) \middle| s_t^i\right] \tag{4}$$

subject to the following constraints

$$\begin{aligned}
\mathbb{I}\left(\{d_t\}; \{s_t^i\}\right) &\leq \kappa^i && (5)\\
q_{t+1}^{i*} &= \arg\max_{q_{t+1}^i} \mathbb{E}\left[u(c_t^i; s_t^i) + \beta u(c_{t+1}^i; s_t^i) \middle| s_t^i\right] && (6)\\
c_t^i &= q_{t+1}^{i*}\left(d_t - p_t\right) && (7)\\
c_{t+1}^i &= q_{t+1}^{i*} p_{t+1} && (8)
\end{aligned}$$

In the first stage (4) the agent decides on the signal structure he wants to receive taking his information processing constraint (5) into account. In the second stage he decides on his trading strategy (6) given the received signal of the chosen structure and the budget constraints (7) and (8).

Note that in this framework agent i decides on the level of precision of the signal $s_t^i \in \Gamma$ and not a specific signal itself. Furthermore, following Sims (2003) and using Shannon entropy as an information measure implies that the precision level of a signal can be translated into the correlation between dividend and signal.

Proposition 1. *Information Capacity Constraint*
Given independent dividends and signals the information capacity constraint (3) can be written as:

$$\underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_1^2}\right)}_{\kappa_1} + \underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_2^2}\right)}_{\kappa_2} \dots . + \underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_N^2}\right)}_{\kappa_N} \leq \kappa \tag{9}$$

Where ρ_n is the chosen correlation parameter between asset n's dividend and the corresponding signal.

Proof: See section VII.A. □

Solving the agent's decision problem one has to start at the second stage, finding the market price dependent on the signals received. Here we assume that the agents are identical and thus each generation can be represented by on singel agent, we relax this assumption the course of this paper. Together with the market clearing condition, one can derive the market price analogously to Lucas Jr (1978). Having a solution for the market price depended on the structure of the signals, the first stage of the problem can be addressed by weighting the usefulness of a more precise signal for one asset against less precise signals for all other assets and thus solving the attention allocation problem.

A. Two Assets and One Agent Group

This section introduces the simplest case of our model in order to get some intuition towards the approach and the equilibrium. The results are similar to the ones when using a Lucas tree model such as Peng (2005) or Luo and Young (2010). Thus, while this part does not add new insights, there are some interesting implications gained on the way of getting there.

We assume the case of an economy with two risky assets, think of them as shares in a company, bonds or a portfolio of financial products, and one group of homogeneous agents with mean-variance utility, see Biais et al. (2010):

$$u(c) = \mathbb{E}\left[c\right] - \frac{\gamma}{2} Var\left[c\right] \tag{10}$$

$$u(c; s) = \mathbb{E}\left[c|s\right] - \frac{\gamma}{2} Var\left[c|s\right] \tag{11}$$

Since we have Baysian-Agents, in equilibrium they weight between their prior information, μ, and the information received by the signal. The weights are determined by how noisy or informative the signals are, thus the more informative, less noisy, a signal is, the more weight the agent will put on this signal. Therefore the signal choice influences the expected value of the asset and thus the market clearing price, as can be seen in equation (12). Also note, that the less informative the signal, the higher will be A and thus the higher will be the risk-premia. Thus a lack of information processing capacities leads to noisy information and higher risk-premia, pressing the prices.

$$p_t \quad = \quad \Xi^2 \cdot (s_t - \mu) + \frac{\mu}{1-\beta} - \frac{\gamma}{(1-\beta)} A1 \tag{12}$$

Ξ is the diagonal matrix of optimal correlations ρ_n^* and A the variance of future dividends and prices:

$$A = \begin{pmatrix} \sigma_1^2(1-\rho_1^{*2}) + \beta\sigma_1^2\rho_1^{*2} & \\ & \sigma_2^2(1-\rho_2^{*2}) + \beta\sigma_2^2\rho_2^{*2} \end{pmatrix} \tag{13}$$

Once the reaction of the market price towards changes in the information choice is clear, the agent can solve the attention allocation, leading to the following solution, proposition (2) is about.

Proposition 2. *Attention allocation in the Two Asset Case*
The optimal attention allocation structure, meaning the desired signal precision,

which depends on the variance ratio of both dividend processes $\varkappa = \frac{\sigma_1}{\sigma_2}$, *as:*

$$\rho_1^* = \begin{cases} \sqrt{1-\left(\frac{1}{4}\right)^\kappa} & if \quad \varkappa^2 > 4^\kappa \\ \sqrt{1-\frac{1}{\varkappa}\left(\frac{1}{2}\right)^\kappa} & if \quad \varkappa^2 \in \left[\frac{1}{4^\kappa};4^\kappa\right] \\ 0 & if \quad \varkappa^2 < \frac{1}{4^\kappa} \end{cases} \tag{14}$$

$$\rho_2^* = \begin{cases} 0 & if \quad \varkappa^2 > 4^\kappa \\ \sqrt{1-\varkappa\left(\frac{1}{2}\right)^\kappa} & if \quad \varkappa^2 \in \left[\frac{1}{4^\kappa};4^\kappa\right] \\ \sqrt{1-\left(\frac{1}{4}\right)^\kappa} & if \quad \varkappa^2 < \frac{1}{4^\kappa} \end{cases} \tag{15}$$

Proof: See Appendix (VII.B) □

Assuming that one asset has a lower variance than the other, proposition 2 states that the agent will allocate more attention on tracking the more risky asset (i.e. the more volatile one), since the gain in utility by eliminating uncertainty is higher with this asset. The larger the differences in the variance and the lower the information processing capacity κ the more available attention is allocated to the asset with the higher variance.[12] Under the information processing constraint allocated attention becomes valuable. The value of the information processing capacity can be measured by the expected excess price the agent group attributes to the asset under the information processing constraint κ as compared to the uninformed state. Lets demonstrate this by having a look at one of the corner solution, where $\varkappa^2 > 4^\kappa$, collapsing to the one asset case.[13] Than ρ^{*2} is directly given by

$$\rho_1^{*2} = 1 - \frac{1}{4^\kappa} := \alpha\left(\kappa(t)\right) \tag{16}$$

Taking expectations of the price in the case without any signal:

$$\mathbb{E}[p] \quad = \quad \frac{\mu - \gamma\sigma^2}{1-\beta} \tag{17}$$

Doing the same in the constrained case with a binding information processing constraint (3):

$$\mathbb{E}[p] \quad = \quad \frac{\mu - \gamma\sigma^2}{1-\beta} + \gamma\left(1 - \frac{1}{4^\kappa}\right)\sigma^2 \tag{18}$$

[12] An numerical solution for the N asset case by an approximation of the resulting boundary conditions will be discussed in section VII.F.

[13] Could also be $\varkappa^2 < \frac{1}{4^\kappa}$.

The excess value is bounded by $\gamma\sigma^2$ and we can define the relative excess value $X(\cdot)$, meaning which part of the maximal excess value is achieved. Defining the excess value as a function of κ leads to:

$$X(\kappa) = \gamma\sigma^2\left(1 - \frac{1}{4^\kappa}\right) \tag{19}$$

For the inner solutions, thus the actual two asset case the excess value of information is:

$$X(\kappa) = 2\gamma\left(\left(1 - \frac{\sigma_2}{\sigma_1}\sqrt{\frac{1}{4^\kappa}}\right)\sigma_1^2 + \left(1 - \frac{\sigma_1}{\sigma_2}\sqrt{\frac{1}{4^\kappa}}\right)\sigma_2^2\right) \tag{20}$$

$$= 2\gamma\left(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\sqrt{\frac{1}{4^\kappa}}\right) \tag{21}$$

The upper bound is given by $2\gamma\left(\sigma_1^2 + \sigma_2^2\right)$. Obviously this excess value is structurally different to the one asset case. On the one hand both variances come naturally into play, but there is also a punishment term, reflecting the trade-off between the attention towards both assets. In the symmetric case $\sigma_1 = \sigma_2$ the excess value becomes:

$$X(\kappa) = 4\gamma\sigma^2\left(1 - \frac{1}{4^\kappa}\right) \tag{22}$$

The value of information processing capacity will get more important later on, when having the possibility to interact with other agent groups and the value can be implemented to actual gains and not only increases in utility reflected by the asset price. For now it is enough to point out the definitive value of information processing capacity within our model.

B. *Heterogeneous Informed Agents*

Allowing for heterogeneity in the information leads to groups of agents, who receive different signals, but are otherwise identical. Which leads to derive first testable empirical hypotheses in the next step.

Proposition 3. *Price Variance*
Assume that there are G equally large groups of agents with independent signals and one asset in the economy, then the price formula is given by:

$$p_t = \rho^{*2}\left(\frac{1}{G}\sum_{g=1}^{G} s_{gt} - \mu\right) + \frac{\mu}{1-\beta} - \frac{\gamma}{(1-\beta)}A1 \tag{23}$$

with

$$A = \sigma^2(1-\rho^{*2}) + \beta\left(\sigma^2 + \frac{1}{G}\tilde{\sigma}^2\right)\rho^{*4} \tag{24}$$

Consequently the variance of the price of the first asset can be written as:

$$Var\left(p_t\right) = \left(\sigma^2 + \frac{1}{G}\tilde{\sigma}^2\right)\rho^{*4} \tag{25}$$

Proof: See section VII.C. □

Proposition 3 directly links the price variance of an asset to the variance of the underlying dividend process, the attention allocated on the asset, and the number of independent signals, meaning the number of different agent groups G.

For heterogeneous informed market participants, meaning a very high G, the volatility of an asset price depends only on the variance of the dividend and the attention allocation, given a not too small attention allocation and thus not a too big signal variance. On the other hand, if the number of heterogeneous informed groups G decreases the variance of the price increases, as long as the information processing capacity stays constant. This is intuitive plausible, since the more different agents are in the market, the likelihood is higher, that new information enters the market, which reduces the likelihood of an shock due to not perceived information.

C. A First Empirical Test to the Model

In a next step we will rewrite our model in a time series regression form in order to fit it to data later on. Assuming a large G, which should be the case when looking at an highly liquid asset or basket of assets, the $\frac{1}{G}\tilde{\sigma}^2$ term becomes insignificantly small:

$$Var\left(p_t\right) = \sigma^2\rho^{*4} \tag{26}$$

To bring the model to the data one has further to assume an exogenously given attention allocation, since there is no possibility to solve the attention allocation problem without knowing κ, the number of possible sources of risk, and their fundamental variance. Taking the root and replacing the correlation of the signal with the fundamental ρ^{*2} by $\alpha\left(\kappa\left(t\right)\right)$, see (16), as well as introducing time dependency leads to the following form of the model:

$$\sigma\left(p_t\right)\left(t\right) = \sigma\alpha\left(\kappa\left(t\right)\right) \tag{27}$$

As our asset we chose the S&P 500 index and a reasonable choice for an exogenously given attention allocation measure on the S&P 500 would be the Google Investing Index (GII), which captures all finance related Google searches in the

US. Since we are not interested in the S&P 500 itself but in its instantaneous volatility, we take the VIX S&P 500 implied volatility index as a proxy. Furthermore we assume $GII(t) \sim \alpha(\kappa(t))$ with a coefficient of proportionality of χ and normalize all values.[14] For our regression we take daily data from September 23rd, 2008 until September 23rd, 2013. This is equal to 1282 observations. The VIX S&P 500 is obtained from Datastream and the Google Investing Index from the Google website.[15]

FIGURE 2. S&P 500 VOLATILITY AND GOOGLE INVESTING INDEX

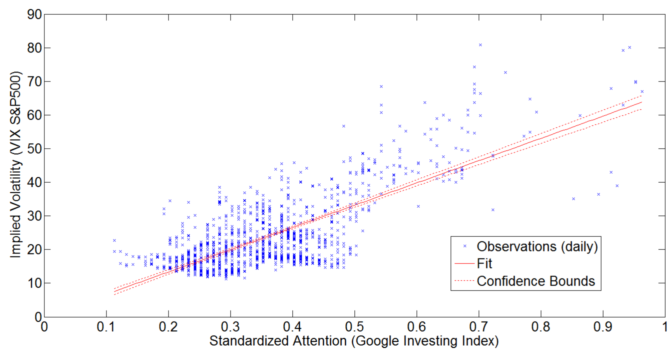


Figure 2 shows the VIX S&P500 implied volatility index (Datastream) and the Google Investing Index (Google Finance), normalized by subtracting 0.38, during the time from September 23rd, 2008 until September 23rd, 2013. The number of observations is 1282. Further the resulting fit of the corresponding linear regression is shown, as well as it's 95% confidence bounds. The results of the regression are portrayed in table 1.

If the attention measured by the Google Investing Index would be irrelevant one would see a symmetric cloud and an insignificant regression coefficient χ. As figure 2 and our regression analysis show this is not the case. We find an adjusted R^2 value of 0.55 and a regression coefficient for the Google Investing Index significant at the 1% level as well as a very high F-statistic of 159 of our model.[16] Looking at these results the first question, which comes to mind, is of course causality. Since the Google Investing Index is only available as a seven day's average, it consists mostly of data prior to the VIX S&P 500 volatility index. Normally the next step towards determining causality would be a test of Granger causality, which unfortunately is not feasible in this particular case, since the

[14]The normalization is performed by subtracting 0.38, which makes the intercept in the later on performed regression approximately zero. Since we not know the absolute level of attention at any time and only assume relative changes, setting the intercept to zero is a valid normalization choice.

[15]http://www.google.com/finance?cid=2001035, viewed September 23rd, 2013

[16]The results of this regression should be interpreted with care, since the underlying time series are non-stationary, which might lead to a spurious regression analysis. Given the long-term mean reversion character of volatility and the Google Investing Index it should however not be a critical issue in this case.

underlying time series are non-stationary and the Google Investing Index's first difference process is quasi discrete. However, Da et al. (2011) show using Google search data on Bloomberg stock ticker numbers that Google search volume leads other attention measures, such as extreme returns or news. Other studies indicate in the same direction, e.g. Preis et al. (2013) provide evidence for the predictive power of changes in Google search volume and Moat et al. (2013) show the numbers of readers of Wikipedia articles related to financial topics are "early warning signs" for stock markets moves. Thus it is quiet save to assume causality to go the way from attention to volatility and not the other way around.

Table 1—S&P 500 Volatility and Google Investing Index

$\sigma\,(S\&P500)\,(t)$	Intercept	χ	adj. R^2
$\sim GII\,(t)$	0.09	66.17***	0.55

This table shows the results of the time series regression of the VIX S&P 500 implied volatility index on the normalised Google Investing Index. The observation period is consisting of 1282 observations from September 23rd, 2008 until September 23rd, 2013. Statistical significance is indicated by *, **, and *** at the 10%, 5%, and 1% level respectively. The intercept value is a result of the normalisation process and should not be interpreted in this regression. The corresponding F-statistic of the model against the constant model is 159.

There are three major implications towards how information travels in financial markets and how efficiently it is incorporated into the price from these findings. First (1), the information flow, which is represented by the signals in our model, appears to be limited by the capacity to process information and by how much attention or information processing capacity is spent on a particular source of uncertainty measured by the Google Investing Index. Second (2), this clearly contradicts the efficient market hypothesis, at least in the semi-strong and strong form, see Fama (1970). According to the efficient market hypothesis all publicly available information should always be rapidly incorporated into the price and thus there should be no relationship between Google search volume and volatility. It appears however that the attention to the information is relevant, not just its availability. Finally (3), information is incorporated efficiently into the price if the allocated information processing capacity is sufficiently high. Consequently, the market is supposed to be efficient during repeated events, which make information available like for example earning announcements. In sum this implies that testing for market efficiency in an event study context as pioneered by Fama et al. (1969) is problematic, since one is testing market efficiency of processing information conditional on the level of attention allocation. Thus, rephrasing it, such an event study picks situations, during which information is better processed than in normal times, since attention allocation is high during such events, in order to proof that information is incorporated into the asset prices by financial markets efficiently in general.

III. One Asset with Shift

The last section dealt with how much available information is actually incorporated into the price and how this depends on the information processing capacity allocated to this task. The natural follow-up question is on how does information travel into the market price. To answer this question we use the inter-temporal nature of our framework by introducing shifts in fundamentals of the dividend process and fanalyzing when they are reflected in asset prices. Specifically, we look at the behaviour of an asset from the end of one equilibrium state given by an exogenous shock in the long-term mean of the dividend process till the next shock. Thus we augment our model by a shift $\tilde{\mu}$ in the fundamental dividend process, which is distributed normally with mean μ_0 and variance σ_2^2. The agent receives two signals, one about the dividend itself s_{1t} and one about the shift in the mean s_{2t}. Thus the dividend and signal processes look as follows:

$$d_{1t} = \mu + \sigma_1 \epsilon_{1t} + \tilde{\mu} \tag{28}$$

$$\tilde{\mu} = N(\mu_0, \sigma_2^2) \tag{29}$$

$$s_{1t} = \mu_1 + \sigma_1 \epsilon_{1t} + \tilde{\sigma}_1 \psi_{1t} \tag{30}$$

$$s_{2t} = \tilde{\mu} + \tilde{\sigma}_2 \psi_{2t} \tag{31}$$

All noise terms are independent of each other and the correlations, ρ_1 and ρ_2, of d_1 with s_{1t} and $\tilde{\mu}$ with s_{2t} depend only on the variance of the additional noise terms. The best guess of the agent about $\tilde{\mu}$ is denoted with μ_t, which is the weighted mean between the agent's signal on the mean s_{2t} and his previous guess μ_{t-1}, thus $\mu_t = \rho_2^2 s_{2t} + (1 - \rho_2^2)\mu_{t-1}$. To ensure a stationary problem and since the agent does not know if a shift has occurred every generation assumes its information about $\tilde{\mu}$ to have variance σ_2^2, meaning the same quality.

Proposition 4. *Attention Allocation with Shift*
If there is only one asset with a shift in the mean of the dividend process and one agent group with mean-variance utility, the price for the asset is given by:

$$p_t = \rho_1^{*2}(s_{1t} - \mu) + \frac{\rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t}{1 - \beta} - \frac{\gamma}{(1-\beta)} A1 \tag{32}$$

The variance is given by:

$$A = \sigma_1^2(1 - \rho_1^{*2}) + \frac{\sigma_2^2}{(1-\beta)^2}(1 - \rho_2^{*2}) + \beta\sigma_1^2\rho_1^{*2} + \frac{\beta}{(1-\beta)^2}\rho_2^{*2}\sigma_2^2 \tag{33}$$

This implies the attention allocation:

$$\rho_1^* = \begin{cases} \sqrt{1-\left(\frac{1}{4}\right)^{\kappa}} & if \quad \varkappa^2 > 4^{\kappa} \\ \sqrt{1-\frac{1}{\varkappa}\left(\frac{1}{2}\right)^{\kappa}} & if \quad \varkappa^2 \in \left[\frac{1}{4^{\kappa}};4^{\kappa}\right] \\ 0 & if \quad \varkappa^2 < \frac{1}{4^{\kappa}} \end{cases} \tag{34}$$

$$\rho_2^* = \begin{cases} 0 & if \quad \varkappa^2 > 4^{\kappa} \\ \sqrt{1-\varkappa\left(\frac{1}{2}\right)^{\kappa}} & if \quad \varkappa^2 \in \left[\frac{1}{4^{\kappa}};4^{\kappa}\right] \\ \sqrt{1-\left(\frac{1}{4}\right)^{\kappa}} & if \quad \varkappa^2 < \frac{1}{4^{\kappa}} \end{cases} \tag{35}$$

with $\varkappa = \frac{\sigma_1}{\sigma_2}(1-\beta)$.

Proof: See section VII.D. □

The main difference to the model without a shift is that the attention allocation does not only depend on the signal to noise ratio any more. The time preference also influences the attention allocation on the long-term dividend mean or on the dividend today. The less the agent is concerned about tomorrow, meaning the lower the β, the less he focuses on the mean shift. At this point it needs to be stressed that a constant β over all generations is assumed.

Proposition 5. *Intertemporal Attention Allocation*
The attention allocation varies with the time preferences β. The higher β the lower will be the attention on the short-term component ρ_1^* and the higher will be the attention on the long-term component ρ_2^*.

Proof: The proposition follows immediately from $\varkappa = \frac{\sigma_1}{\sigma_2}(1-\beta)$ and its influence on the allocation scheme. □

A. *Simulation and Empirical Evidence of the Shift Model*

The partial neglect of the fundamentals leads to a lacked adjustment towards the new equilibrium. Figure 3 shows one simulated sample price path, where the shift, $\tilde{\mu} = -2$, in period T is considerably large. The simulation was performed in Matlab using $\beta = 0.9$, $\gamma = 0.2$, $\kappa = 0.09$, $\mu = 25$, $\mu_0 = 0$, $\sigma_1^2 = 10$, and $\sigma_2^2 = 0.1$ as the underlying market parameters. Since $\frac{\sigma_1}{\sigma_2}(1-\beta) = 1$ the attention is equally distributed in this case. For the sake of simplicity we will choose the model specification always in such a manner, that attention is allocated equally between the long- and short-term sources of uncertainty when performing simulations.

One can see that the adjustment time of the price shown in figure 3 is considerably long. Obviously the two factors most important for determining the adjustment time are the size of the shift and the information processing capacity κ. To give an overview of the interdependencies of these three variables we

Figure 3. Sample Path of a Shift

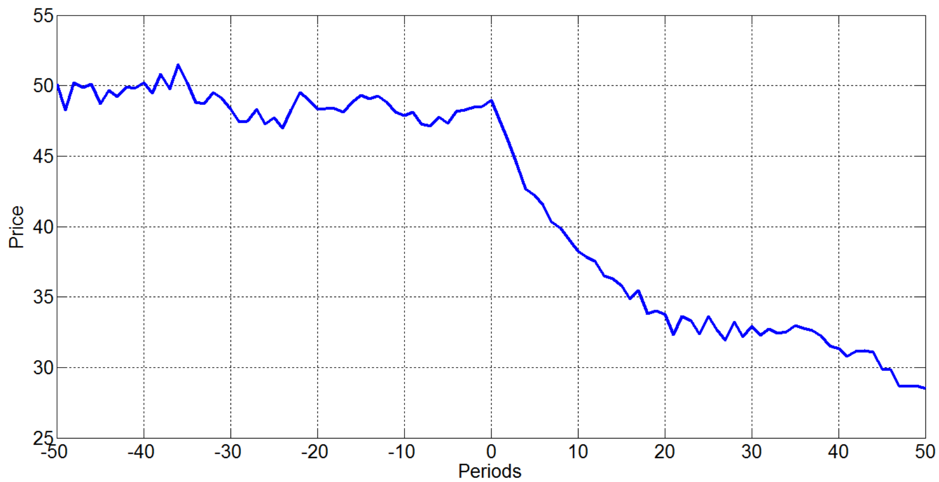


Figure 3 shows one sample path of a simulated asset price during a shift. The period of the shift is $t = 0$. The whole graph was simulated with $T = 0$, $\beta = 0.9$, $\gamma = 0.2$, $\kappa = 0.09$, $\mu = 25$, $\mu_0 = 0$, $\sigma_1^2 = 10$, $\sigma_2^2 = 0.1$, and $\tilde{\mu} = -2$ as the underlying market parameters.[17]

perform a sensitivity analysis of the adjustment time as a function of the information processing capacity and the shock in the long-term mean of the underlying dividend process. Figure VI.A in the appendix shows this analysis for $\kappa \in \{0.10, 0.11, ..., 0.20\}$ and $\tilde{\mu} \in \{1.0, 1.5, ..., 6.0\}$, measured in standard deviations σ_2. The adjustment time is given by the mean adjustment time over 10000 Monte Carlo simulations at each node and defined as the time it takes the price to reach its new theoretical long-term mean for the first time after the shift occurred. Looking at figure VI.A one can see an overproportional increase in the adjustment time with lower information capacity κ and an underproportional increase with shift size.

Having illustrated the properties of our model by two simulation studies we turn now to an empirical example. Perhaps the best example data-wise of our model is the burst of the US subprime bubble, since it represents a major shift in an asset price during a period for which Google search data is available. Figure 4 shows the Case-Shiller Home Price 20 City Composite index and the cumulative Google Trend search results of "subprime" in the period of January 2006 to December 2011. The Case-Shiller Home Price 20 City Composite is an index of the home prices of the 20 major metropolitan areas in the US. The index is published monthly by Standard & Poor's. It uses the Karl Case and Robert Shiller method of a house price index, which is a modified version of the weighted repeat sales methodology.[18] The cumulative Google Trend search results of “subprime” were directly obtained from Google Trend.[19] Since there exists a base rate of non-financial related searches for “subprime” only each monthly value in excess of the long-term sample mean of 12 in the original measure of Google Trends were used. As one can see, the increase in Google search volume for “subprime” presides

Figure 4. Case-Shiller Home Price Index and Google Trend Search “Subprime”

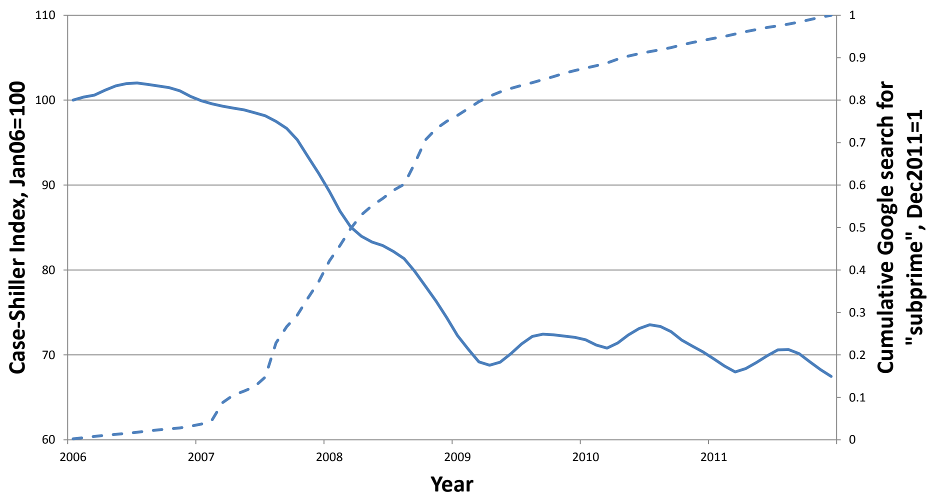


Figure 4 shows the Case-Shiller Home Price 20 City Composite index (left axis) with January 2006 as basis 100 and the cumulative Google Trend search results for “subprime” (right axis) with its December 2011 set to 1. The data of the Case-Shiller index is obtained from Datastream. The cumulative Google Trend search results of “subprime” were directly obtained from Google Trend. Since there exists a base rate of non financial related searches for “subprime” each monthly value in excess of the long-term sample mean of 12 in the original measure of Google Trend was used.

the major downturn in the second half of 2007 in the Case-Shiller Home Price Index. During the price correction more and more available information about the housing market is absorbed by market participants informing themselves, which is measured by the number of cumulative Google searches for this topic. This also implies that traders with an informational advantage could gain a lot out of their obtained information. We will return to this question in section IV.B, where we explicitly model a market of heterogeneous agents with respect to their information processing capacity constraint during a shift. Going back to the question of how efficient information is processed by the market, the burst of the subprime bubble is a good example for this not being the case and how it takes a long time for the information to spread as more and more market participants

[18]http://www.spindices.com/index-family/real-estate/sp-case-shiller, viewed September 25th, 2013
[19]http://www.google.de/trends/, viewed September 23rd, 2013

inform themselves about the underlying situation. It is important to point out again that only singular events can provide evidence toward market inefficiencies as discussed in section II.B.

B. *Implications for the Momentum Effect*

As the simulation in figure (3) shows, the model generate phases of uniformly positive or negative expected earnings in excess of dividends, which can be seen as excess returns and are empirically resembled by a phase of autocorrelation. This pattern is a necessary requirement for a momentum effect (Biais et al., 2010). Figure (5) elaborates this notion of correlated excess returns during a shift in a simulation of our model. The autocorrelation of the full sample is close to zero. But during the transition phase from period 100 till period 125 the autocorrelation is substantially higher with 0.24.

FIGURE 5. MOMENTUM TRADING

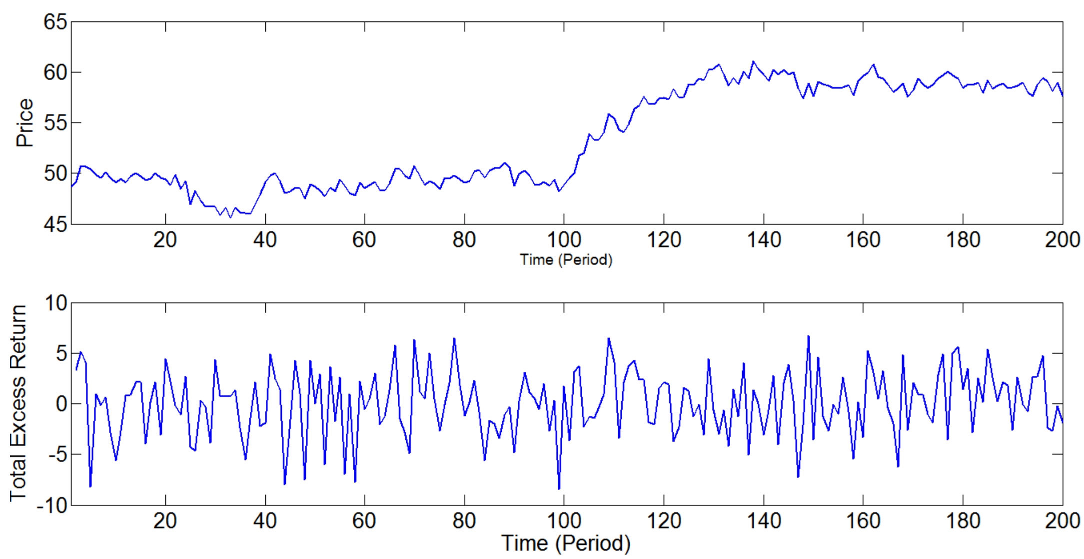


Figure 5 shows the asset sample path, with a positive shift in the long-term mean of the dividend process in period 100. The graph below shows the excess returns, meaning the returns in excess of dividends. $\beta = 0.9$, $\gamma = 0.2$, $\kappa_1 = 0.20$, $\kappa_2 = 0.25$, $\mu = 10$, $\sigma_1^2 = 5$, $\sigma_2^2 = 0.05$, and $\tilde{\mu} = 2$ are the parameters of the underlying market. The autocorrelation of the excess returns over the full sample is -0.08 and 0.24 during the shift from period 100 till period 125.

Proposition 6. *Momentum Effect*
During a shift, meaning a phase of length $\tau \in \mathbb{N}$ *with* $\tilde{\mu} < \mu_{t+i}$ *or* $\tilde{\mu} > \mu_{t+i}$ $\forall i \leq \tau$*,*

the expected excess return under an extended outsider information set[20] *is either uniformly positive or negative.*

Proof: See section VII.E. □

The intuition is that during a positive (negative) shift, the expected excess return conditional on the shift, meaning if the agent knows that there has been a shift, is also positive (negative). This means that there is a statistical autocorrelation if one looks ex-post at the data, but ex-ante, in the moment the agent has to decide how to invest, the agent cannot be sure about the shift and its size since its only source of information is the signal he receives. Consequently, the agent is not able to take advantage of this momentum trading opportunity on a single asset.[21]

IV. Investors with Different Information Processing Capabilities

On the financial market the participants differ in their infrastructure and expertise on interpreting available data of a financial services provider, particularly when one compares investment banks with an amateur investor. In this view the financial services industry is a seller of information processing capacity when offering products to non-professional investors. Another example is from within professional trading, where computer models and algorithms are taking over more and more human decision making. For example high-frequency traders are involved in almost 70% of all dollar volume trades, see Brogaard (2010). The reasons for this development are simple, the hard- and software used for high frequency trading can process more information in a shorter time as compared to a human. Such advantages in information processing leads to higher precision of the information or to a time advantage. Within our framework we can discuss these processing advantages by allowing for different information processing capacity constraint and show how this affects the optimal asset allocation and the returns for different agent groups.

A. Two Assets and Two Types of Agents

We add to the simplest case of a two assets economy as described in section II.A two possible types of agents differing from each other by their information processing capabilities.

Proposition 7. *Let the fraction of group one with information processing constraint* κ^1 *be* λ. *Its information allocation will be denoted by* ρ_{11} *and* ρ_{12}. *The*

[20] The extended outside information set includes the information that a shift is happening, which of course is no information the agent could obtain from only taking into account the signals he receives.

[21] Leaving the scope of our model and assuming an economy with many assets, which are partly shifting at any given time, one would be able to exploit these autocorrelations with an momentum trading strategy, even though one does not know if any particular asset is really shifting or not.

same applies for group two with an information processing constraint κ^2, representing a fraction of the population of $1-\lambda$ and an information allocation of ρ_{21} and ρ_{22}. Then the quantities hold in equilibrium are:

$$q_1^* = \frac{(1-\beta)}{\gamma}\left(A_1+\tilde{A}\right)^{-1}\left(\Xi_1^2\cdot(s_{1t}-\mu)+\frac{\mu}{1-\beta}-p_t\right) \tag{36}$$

$$q_2^* = \frac{(1-\beta)}{\gamma}\left(A_2+\tilde{A}\right)^{-1}\left(\Xi_2^2\cdot(s_{2t}-\mu)+\frac{\mu}{1-\beta}-p_t\right) \tag{37}$$

A_1 and A_2 are the variances associated with the dividends for each of the investor groups and $\tilde{A}$ the variances of future prices. The price of the assets are thus given by:

$$p_t = \frac{\mu}{1-\beta} + \omega\left(\Xi_1^2\cdot(s_{1t}-\mu)\right) + (1-\omega)\left(\Xi_2^2\cdot(s_{2t}-\mu)\right) - \Omega 1 \tag{38}$$

with

$$\omega = \left(\lambda\tau_1 + (1-\lambda)\,\tau_2\right)^{-1}\lambda\tau_1 \tag{39}$$

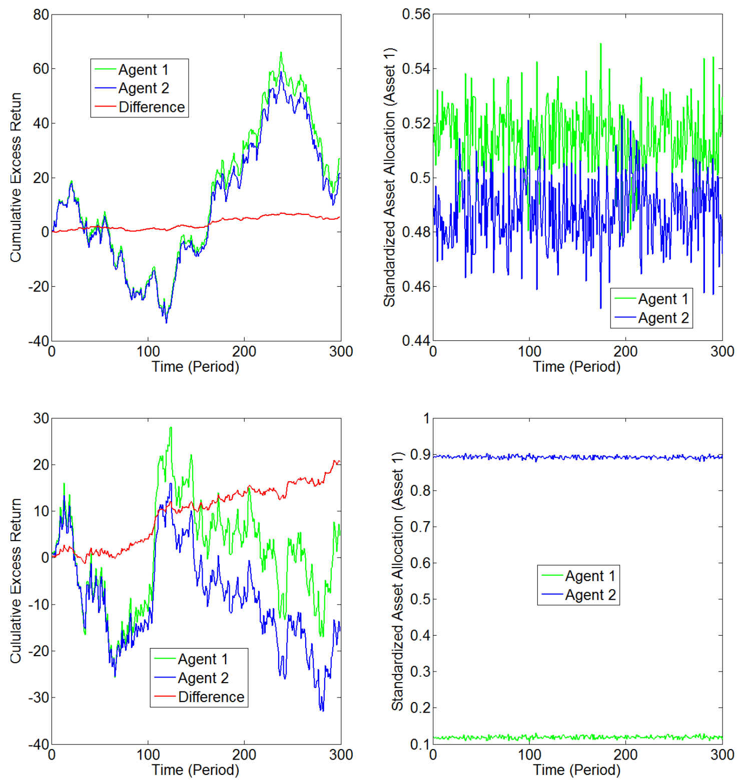
$$\Omega = \left(\lambda\tau_1 + (1-\lambda)\,\tau_2\right)^{-1} \tag{40}$$

Proof: See Appendix VII.G for detailed derivation.

Economically ω represents the fact that λ needs to be adjusted for the average informativeness of the signal of the agent group, since a more precise informed group will trade more in the market as compared to a less informed group of the same size. Figure 6 shows the asset allocation and the cumulative excess returns, meaning the cumulative returns in excess of dividends normalised by group size, for each of the two agent group with different information processing capacities. The simulation parameters of figure 6 differ only in the relative size λ of agent group one, which has a higher information processing capacity. This shows that for the group with a higher information processing capacity constraint the value of κ depends on their own fraction on the whole population of agents. Thus the smaller the group of fast learners is the higher is their excess return (standardized by the group size). This simply reflects the fact, that the group with better information processing capacity constraints faces less competition on information the smaller they are, thus the higher will be the margin for each individual trader.
Due to the more complex and semi-analytical character of the heterogeneous model on the one hand and the different character of the money equivalent of the information processing capacity it is not possible to quantify the value of information processing capacity in a similar manner as in the baseline model, see equations (19) and (22). Specifically, in the base line model the implicit ex-ante gain in utility was presented while in the heterogeneous model ex-post excess

returns are analyzed.

FIGURE 6. VALUE OF INFORMATION PROCESSING CAPACITY



This figure shows the cumulative excess return standardized by group size and the asset allocation of the first asset of two equally big agent groups (upper row, $\lambda = 0.5$) and of two differently big agent groups (lower row, $\lambda = 0.1$) with different information processing capacities trading on two assets. The parameters of the underlying market are taken as $\beta = 0.9$, $\gamma = 0.5$, $\lambda = 0.5$, $\kappa^1 = 0.45$, $\kappa^2 = 0.25$, $\mu_{1,2} = 10$, $\sigma^2_{1,2} = 4$. The length of the simulation is 300 time periods.

B. One Asset with Shift and Two Types of Agents

Instead of having two asset to concentrate on, one can also ask the question regarding two different kind of shocks on the same asset, as we did before in section (III) and exploied the intertemporal framework to discuss such coups like the successful bet of John Paulson on the burst of the subprime bubble of the US housing market from an informational processing and attention allocation point of view. Abstracting from the individual case one can interpret it as an informational advantage from a higher information processing constraint within the environment of a shock in the fundamentals. Model-wise we add to our model of section III two groups of agents with different information processing constraints.

Proposition 8. *In addition to proposition (7), each group of agents has now current beliefs μ_{1t}, μ_{2t} of the long-term mean, which they update every period. This allows for heterogeneity in the dynamics in the long-term mean expectations. The optimal asset allocation of both groups is given by:*

$$q_1^* = \frac{(1-\beta)}{\gamma}\left(A_1+\tilde{A}\right)^{-1} \left(\rho_1^{*2}\left(s_{11t}-\mu\right)+\frac{\rho_2^{*2}\left(s_{12t}-\mu_t\right)+\mu+\mu_{1t}}{1-\beta}-p_t\right) \tag{41}$$

$$q_2^* = \frac{(1-\beta)}{\gamma}\left(A_2+\tilde{A}\right)^{-1} \left(\rho_1^{*2}\left(s_{21t}-\mu\right)+\frac{\rho_2^{*2}\left(s_{22t}-\mu_t\right)+\mu+\mu_{2t}}{1-\beta}-p_t\right) \tag{42}$$

Using the same notational short cuts as before the price will be:

$$\begin{aligned} p_t = {} & \omega\left(\rho_1^{*2}\left(s_{11t}-\mu\right)+\frac{\rho_2^{*2}\left(s_{12t}-\mu_t\right)+\mu+\mu_{1t}}{1-\beta}\right) \\ & +(1-\omega)\left(\rho_1^{*2}\left(s_{21t}-\mu\right)+\frac{\rho_2^{*2}\left(s_{22t}-\mu_t\right)+\mu+\mu_{2t}}{1-\beta}\right) \\ & -\Omega_1 \end{aligned} \tag{43}$$

with

$$\omega = \left(\lambda\tau_1+(1-\lambda)\,\tau_2\right)^{-1}\lambda\tau_1 \tag{44}$$

$$\Omega = \left(\lambda\tau_1+(1-\lambda)\,\tau_2\right)^{-1} \tag{45}$$

Proof: For detailed derivation see Appendix (VII.H).

Having derived the model we now want to compare its implications, which we portray with the help of a simulation case study, with the bet of John Paulson on the burst of the US housing market bubble. Figure 7 shows a simulation of the most extreme case, meaning a small group with a high information processing capacity and a large group with a very small information processing capacity constraint. As one can see, both groups trade in opposite directions as most of the future value of the asset is wiped out by a shock in period 5. While the group with a high information processing capacity constraint shortens the asset, the other one is buying it, since it is not yet aware of the sharp drop.
By shorting his exposure to BBB tranches of subprime mortgage backed securities using credit default swaps John Paulson was able to reap a huge profit when the

US housing market declined sharply in 2007/2008 and many of the BBB tranches lost all their value. That this was not just a lucky guess, but actually a model example of an agent having a huge advantage in information processing capacity can be seen by the fact that “John Paulson [...] purchased the best database on house-price statistics, commissioned a technology company to help him warehouse it, and hired extra analysts to interpret the numbers”.[22] Wanting to trade on what he perceived as the greatest weakness of the US financial system he allocated a lot of information processing capacity on the US housing market, as one can see from his investments, and was therefore able to exploit the informational advantage during the burst of the bubble. Even though he earned more than $15 billions with this strategy in the end, he lost millions of dollars on the way, since it took a while till the prices dropped finally in 2007/2008.[23]

[22] Mallaby (2010), p. 386
[23] For an in-depth discussion of John Paulson’s trade see Mallaby (2010), p. 307-391.

FIGURE 7. HETEROGENEOUS INFORMATION PROCESSING CAPACITIES AND SHIFT

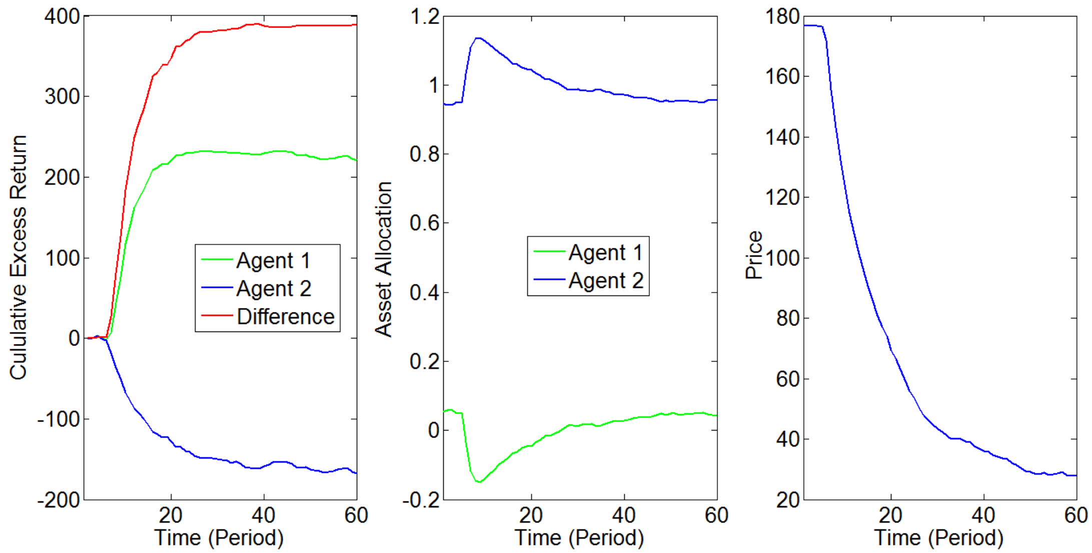


Figure 7 shows the cumulative excess return and the asset allocation, standardized by group size, of each agent group and the dynamics of the price of the shifted asset. The simulation time is 60 periods with a shock in period 5, which wipes out most of the assets fundamental value. The model parameters used for the simulation are $\beta = 0.9$, $\gamma = 0.2$, $\lambda = 0.05$, $\kappa^1 = 0.5$, $\kappa^2 = 0.01$, $\mu = 20$, $\mu_{0,1} = 0$, $\mu_{0,2} = 0$, $\sigma_1^2 = 4$, $\sigma_2^2 = 0.04$, $\tilde{\mu} = -15$.

V. Conclusion

In this paper we introduced a model of rational inattention with an overlapping generation model for the financial market. To the best of our knowledge we are the first to combine rational inattention with a real market model and to derive implications towards competitive attention allocation and the choice between short-term and long-term relevant information from it. Going on from these mostly theoretical results we derived the following four main implications from our model. First (1), we have shown and empirically tested that the capability to process information and the willingness to allocate this capability towards a specific source of uncertainty is highly relevant in the context of how information travels in the financial market. Thus we challenged the efficient market hypothesis by an alternative framework of attention driven efficiency. Given the idea of attention driven efficiency, we further pointed out that it might be problematic to test market efficiency in general during times of high allocated attention, since the market should be efficient during these times. Second (2), extending our basic model by a shift component, we showed how shifts can be seen as a result of limited information processing capacity while still staying in a rational agent framework. Furthermore, we portrayed the plausibility of this concept with an empirical case study of the burst of the US subprime bubble and John Paulson's successful bet on it. Third (3), within this shift framework, we were able to give a micro-level explanation of the momentum effect in a rational agent framework without direct arbitrage opportunities. Fourth (4), we have shown that within our framework financial services providers can be seen as providers of information processing capacity.

Since this is the first rational inattention model developed to explicitly model information aggregation on the financial market, we believe that there is still a huge potential for other possible applications. Further, there should be ample opportunity to extend and improve on the suggested model. Perhaps most prominently the question of making information available to other parties is not addressed within our model context and would be the most interesting extension of our framework.

VI. Appendix

A. *Figure on Adjustment Time Depended on κ and the Shift Size*

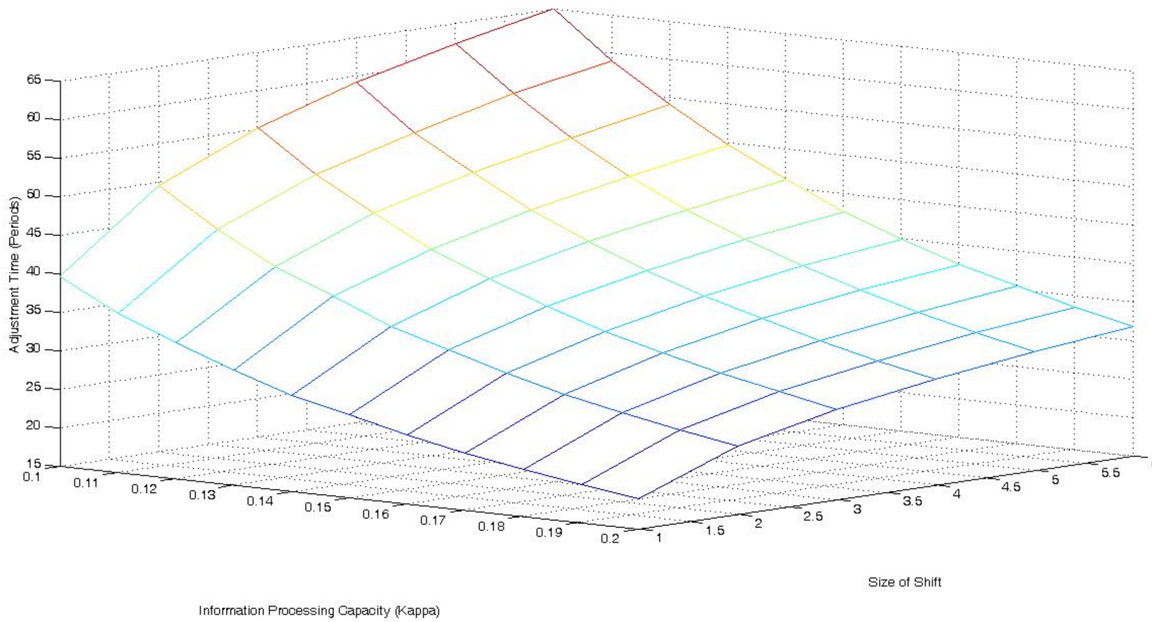


Figure VI.A shows the mean adjustment time, meaning the time starting from when the shock happened until the price reaches the new theoretically implied long-term mean level for the first time. The graph presents the mean adjustment time of 10000 Monte Carlo simulations at each node using $T = 0$, $\beta = 0.9$, $\gamma = 0.5$, $\mu = 20$, $\mu_0 = 0$, $\sigma_1^2 = 10$, and $\sigma_2^2 = 0.1$ as the underlying market parameters.

VII. Mathematical Proofs

A. Proof of Proposition 1

If two normally distributed random variables X and Y are correlated with each other with correlation parameter ρ the mutual information, meaning the information one variable contains about the other, can be expressed as the reduction in the entropy of X by observing the other random variable Y.

$$I(X,Y) = H(X) - H(X|Y) \tag{46}$$

$H(X)$ is the unconditional entropy of X and $H(X|Y)$ is the conditional entropy of the X given Y. Both can be calculated using the entropy formula:

$$H(X) = \frac{1}{2}\log_2\left[(2\pi e)^T \det \Omega_X\right] \tag{47}$$

$$H\left(X|Y\right) = \frac{1}{2}\log_2\left[(2\pi e)^T \det \Omega_{X|Y}\right] \tag{48}$$

If X, Y are jointly multivariate normal distributed with $Cov(X_i, X_j) = 0 \ \forall i \neq j$, $Cov(Y_i, Y_j) = 0 \ \forall i \neq j$, $Cov(X_i, Y_j) = \rho\sigma_X\sigma_Y \ \forall i = j$, $Cov(X_i, Y_j) = 0 \ \forall i \neq j$, $Var(X_i) = \sigma_X^2$, and $Var(Y_i) = \sigma_Y^2$ the mutual information (46) can also be written in the following way:

$$I(X,Y) = H(X) - H(X|Y) \tag{49}$$

$$= \frac{1}{2}\log_2\left[(2\pi e)^T\sigma_X^{2T}\right] - \frac{1}{2}\log_2\left[(2\pi e)^T\left(\sigma_X^2 - \sigma_X^2\rho^2\right)^T\right] \tag{50}$$

$$= \frac{1}{2}T\log_2\left(\frac{1}{1-\rho^2}\right) \tag{51}$$

Since we are interested in the average information per period, we divide (51) by T:

$$\mathbb{I}(X_t, Y_t) = \frac{1}{T}\frac{1}{2}T\log_2\left(\frac{1}{1-\rho^2}\right) \tag{52}$$

$$= \frac{1}{2}\log_2\left(\frac{1}{1-\rho^2}\right) \tag{53}$$

In our case the dividends are independent, as well as the signals. Therefore one can think of independent information processes for each asset and simply take the sum of the amount of information in the given period:

$$\underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_1^2}\right)}_{\kappa_1} + \underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_2^2}\right)}_{\kappa_2}.... + \underbrace{\frac{1}{2}\log_2\left(\frac{1}{1-\rho_N^2}\right)}_{\kappa_N} \leq \kappa \tag{54}$$

□

B. Proof of Proposition 2

Proof: The solution to this problem is derived in four steps. The first step derives the asset demand, the second step solves for the market clearing price, the third step simplifies the attention allocation problem, and the fourth step finally solves the attention allocation problem.

Step 1: The market clearing price can be derived from the solution of (6) given an optimal attention allocation and the received signals s_t. The corresponding FOC (first order condition) is given by:

$$0 = \mathbb{E}\left[d_t|s_t\right] + \beta\mathbb{E}\left[p_{t+1}|s_t\right] - p_t - \gamma A q^* \tag{55}$$

$$0 = \Xi^2 \cdot (s_t - \mu) + \mu + \beta\mathbb{E}\left[p_{t+1}|s_t\right] - p_t - \gamma A q^* \tag{56}$$

This is a stationary problem and all the future periods will be the same in expectations given today's signal:

$$\mathbb{E}\left[p_{t+1}|s_t\right] = \mathbb{E}\left[d_{t+1}|s_t\right] + \beta\mathbb{E}\left[p_{t+2}|s_t\right] - \gamma A q^* \tag{57}$$

Iteratively substituting equation (57) in equation (55) leads to the following representation of the problem:

$$\begin{aligned} 0 = {} & \Xi^2 \cdot (s_t - \mu) + \mu + \sum_{t=1}^{T} \beta^t \left(\mu - \gamma A q^*\right) \\ & + \beta^{T+1}\left(\mathbb{E}\left[p_{t+T+2}\right] \gamma A q^*\right) - p_t - \gamma A q^* \end{aligned} \tag{58}$$

This can be simplified by taking the limit of $T \to \infty$:

$$\lim_{T\to\infty} \sum_{t=0}^{T} \beta^t \left(\mu - \gamma A q^*\right) = \frac{\mu}{1-\beta} + \frac{\gamma}{(1-\beta)} A q^* \tag{59}$$

$$\lim_{T\to\infty} \beta^{T+1}\left(\mathbb{E}\left[p_{t+T+2}\right] \gamma A q^*\right) = 0 \tag{60}$$

Rearranging terms one gets:

$$q^* = \frac{(1-\beta)}{\gamma} A^{-1} \left(\Xi^2 \cdot (s_t - \mu) + \frac{\mu}{1-\beta} - p_t\right) \tag{61}$$

Thus (61) is the optimal trading strategy of the agent group given the attention allocation Ξ and the received signal s_t.

Step 2: In equilibrium the agent group has to hold all assets. Normalizing them to one, $q^* = 1$, yields the equilibrium market prices:

$$p_t = \Xi^2 \cdot (s_t - \mu) + \frac{\mu}{1-\beta} - \frac{\gamma}{(1-\beta)} A1 \tag{62}$$

Step 3: With the market clearing price, the agent can solve the attention allocation problem of the first stage, where he maximizes his utility (I) compared to the case where he receives no signals (II).

The attention allocation problem needs to be viewed from an individual agents point of view, who is competing against all other agents of his generation and the following one.

$$\max_{s_t^i \in \Gamma} \mathbb{E}\left[\underbrace{u(c_t^*; s_t^i) + \beta u(c_{t+1}^*; s_t^i)}_{I} - \underbrace{u\left(c_t^\Delta\right) - \beta u(c_{t+1}^\Delta)}_{II} \middle| s_t^i \right] \tag{63}$$

Since the agent can not influence the second period outcomes by his action,s behaves like a price taker and in equilibrium every agent will hold the same amount of assets, the problem reduces to:

$$\max_{s_t \in \Gamma} \mathbb{E}\left[u(c_t^*; s_t^i) - u(c_t^\Delta) \middle| s_t^i \right] \tag{64}$$

The relevant parameter for the signal choice is only the correlation ρ_n between $d_{n,t}$ and $s_{n,t}$. In the context of our model ρ_n is dependent on $\tilde{\sigma}_n$, the only free parameter, since all noise terms are assumed to be independent. The correlation ρ_n therefore only depends on the variance of the additional noise term of the signal:

$$\begin{aligned}
\rho\left(d_{nt}, s_{nt}\right) &= Cor\left(\mu_n + \sigma_n\epsilon_{nt}, \mu_n + \sigma_n\epsilon_n\tilde{\sigma}_n\psi_n\right) && (65)\\
&= Cor\left(\sigma_n\epsilon_{nt}, \sigma_n\epsilon_n + \tilde{\sigma}_n\psi_n\right) && (66)\\
&= \frac{Cov\left(\sigma_n\epsilon_{nt}, \sigma_n\epsilon_n + \tilde{\sigma}_n\psi_n\right)}{\sqrt{Var\left(\sigma_n\epsilon_{nt}\right) Var\left(\sigma_n\epsilon_n + \tilde{\sigma}_n\psi_n\right)}} && (67)\\
&= \frac{\sigma_n^2}{\sqrt{\sigma_n^2\left(\sigma_n^2 + \tilde{\sigma}_n^2\right)}} && (68)\\
&= \frac{\sigma_n}{\sqrt{(\sigma_n^2 + \tilde{\sigma}_n^2)}} && (69)\\
&= \rho_n \in]0, 1] && (70)
\end{aligned}$$

Consequently, the signal structure can be seen as only dependent on the correlation parameters ρ_1 and ρ_2.[24] Combining this with the fact that in equilibrium

[24] Since for a corner solution, meaning $\rho_n = 0$, the signal is irrelevant, we can simply set $\tilde{\sigma}_n^2 = \eta$ for any $\eta \in \mathbb{N}$ to close the set.

the expected utility without any informative signal $E\left[u(c_t^{\Delta})\right]$ is constant for all choices of correlation one can reduce the optimization problem regarding the signals to:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \mathbb{E}\left[u(d_t;s_t)|s_t\left(\rho_1,\rho_2\right)\right] \tag{71}$$

Assuming mean-variance utility it follows that:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \mathbb{E}\left[\mathbb{E}\left[d_t|s_t\left(\rho_1,\rho_2\right)\right]-\frac{\gamma}{2}Var\left[d_t|s_t\right]\Big|s_t\left(\rho_1,\rho_2\right)\right] \tag{72}$$

With two assets the aggregated dividend is $d_t = d_{1,t}+d_{2,t}$ leading to the following representation of our optimization problem:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \mathbb{E}\left[\mathbb{E}\left[d_{1t}+d_{2t}|s_t\right]-\frac{\gamma}{2}Var\left[d_{1t}+d_{2t}|s_t\right]\Big|s_t\left(\rho_1,\rho_2\right)\right] \tag{73}$$

This can be decomposed to:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \mathbb{E}\left[d_{1t}|s_t\left(\rho_1,\rho_2\right)\right]+\mathbb{E}\left[d_{2t}|s_t\left(\rho_1,\rho_2\right)\right]-\frac{\gamma}{2}Var\left[d_{1t}+d_{2t}|s_t\left(\rho_1,\rho_2\right)\right] \tag{74}$$

As we are looking for the optimal correlation of the signals and the fundamentals, all constant parameters or level variables can be neglected for the optimization problem and the only uncertainty arises in the second moments. Furthermore, both dividend processes are uncorrelated and thus an equivalent optimizing problem is given by:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} -Var\left[d_{1t}|s_t\left(\rho_1,\rho_2\right)\right]-Var\left[d_{2t}|s_t\left(\rho_1,\rho_2\right)\right] \tag{75}$$

Applying the rules for dependent mean and variance of correlated normally distributed variables:[25]

$$\max_{\{\rho_1\rho_2\}\in[0,1]} -\sigma_1^2\left(1-\rho_1^2\right)-\sigma_2^2\left(1-\rho_2^2\right) \tag{76}$$

Again ignoring all constant parameters for the optimization, the reduced form is:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \sigma_1^2\rho_1^2+\sigma_2^2\rho_2^2 \tag{77}$$

Step 4: For tractability we replace $\left\{\rho_1^2,\rho_2^2\right\}$ by $\{\xi_1,\xi_2\}$ in the reduced opti-

[25] $(x_1|x_2=a)\sim N\left(\hat{\mu},\hat{\Sigma}\right)$, with $\hat{\mu}=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}\left(a-\mu_2\right)$ and $\hat{\Sigma}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

mization problem:

$$\max_{\{\xi_1,\xi_2\}\in[0,1]} \sigma_1^2\xi_1 + \sigma_2^2\xi_2 \tag{78}$$

subject to the information processing constraint

$$\frac{1}{2}\log_2\left(\frac{1}{1-\xi_1}\right) + \frac{1}{2}\log_2\left(\frac{1}{1-\xi_2}\right) \leq \kappa \tag{79}$$

Since the objective function is increasing in both choice variables the information processing constraint will be binding at any maximum. Therefore one can rewrite this constraint imposing strict equality. Reformulating the $\log_2$'s to ln's we arrive at:

$$\frac{\ln(\frac{1}{1-\xi_1})}{\ln(2)} + \frac{\ln(\frac{1}{1-\xi_2})}{\ln(2)} = 2\kappa \tag{80}$$

$$\exp\left(2\kappa\ln(2)\right) = \frac{1}{1-\xi_1}\frac{1}{1-\xi_2} \tag{81}$$

$$\left(1-\xi_1\right)\left(1-\xi_2\right) = \frac{1}{\exp\left(2\kappa\ln(2)\right)} \tag{82}$$

$$\xi_1 + \xi_2 - \xi_1\xi_2 = \underbrace{1 - \frac{1}{\exp\left(2\kappa\ln(2)\right)}}_{\alpha} \tag{83}$$

$$\xi_1 + \xi_2 - \xi_1\xi_2 - \alpha = 0 \tag{84}$$

The first order condition of the corresponding Lagrange auxiliary function

$$\mathcal{L} = \sigma_1^2\xi_1 + \sigma_2^2\xi_2 + \lambda\left(\alpha - \xi_1 - \xi_2 + \xi_1\xi_2\right) \tag{85}$$

are:
$[\xi_1 :]$

$$\sigma_1^2 + \lambda\left(\xi_2 - 1\right) = 0 \tag{86}$$

$[\xi_2 :]$

$$\sigma_2^2 + \lambda\left(\xi_1 - 1\right) = 0 \tag{87}$$

$[\lambda :]$

$$\alpha - \xi_1 - \xi_2 + \xi_1\xi_2 = 0 \tag{88}$$

Dividing (86) by (87)

$$\underbrace{\frac{\sigma_1^2}{\sigma_2^2}}_{\varkappa^2} = \frac{1-\xi_2}{1-\xi_1} \tag{89}$$

The inner solution follows from a reformulation of (88) leading to the trajectory:

$$\xi_2 = \frac{\alpha - \xi_1}{1-\xi_1} \tag{90}$$

Plugging (90) into (89) yields:

$$\varkappa^2 = \frac{1 - \frac{\alpha-\xi_1}{1-\xi_1}}{1-\xi_1} \tag{91}$$

$$= \frac{1-\alpha}{(1-\xi_1)^2} \tag{92}$$

Solving for $(1-\xi_1)$:

$$(1-\xi_1)^2 = \frac{1-\alpha}{\varkappa^2} \tag{93}$$

$$1-\xi_1 = \pm\sqrt{1-\alpha}\frac{1}{\varkappa} \tag{94}$$

Since $\{\xi_1, \xi_2\} \in [0,1]$, the only plausible solution is given by:

$$\xi_1^* = 1 - \sqrt{1-\alpha}\frac{1}{\varkappa} \tag{95}$$

From (89) one can derive the condition for a corner solution, meaning of a state, in which the second asset will be neglected, specifically $\xi_1^* = \alpha = 1 - \left(\frac{1}{4}\right)^{\kappa}$ and $\xi_2^* = 0$, as:

$$4^{\kappa} < \varkappa^2 \tag{96}$$

Using $\kappa_1 = \frac{1}{2}\log_2\left(\frac{1}{1-\xi_1}\right)$ it follows that:

$$\kappa_1 = \begin{cases} \kappa & if \quad \varkappa^2 \quad > \quad 4^{\kappa} \\ \frac{1}{2}\kappa + \frac{1}{2}\log_2(\varkappa) & if \quad \varkappa^2 \quad \in \quad \left[\frac{1}{4^{\kappa}}; 4^{\kappa}\right] \\ 0 & if \quad \varkappa^2 \quad < \quad \frac{1}{4^{\kappa}} \end{cases} \tag{97}$$

Thus after reconverting ξ_1 into ρ_1:

$$\rho_1^* = \begin{cases} \sqrt{1-\left(\frac{1}{4}\right)^\kappa} & if \quad \varkappa^2 > 4^\kappa \\ \sqrt{1-\frac{1}{\varkappa}\left(\frac{1}{2}\right)^\kappa} & if \quad \varkappa^2 \in \left[\frac{1}{4^\kappa};4^\kappa\right] \\ 0 & if \quad \varkappa^2 < \frac{1}{4^\kappa} \end{cases} \tag{98}$$

and due to symmetry:

$$\rho_2^* = \begin{cases} 0 & if \quad \varkappa^2 > 4^\kappa \\ \sqrt{1-\varkappa\left(\frac{1}{2}\right)^\kappa} & if \quad \varkappa^2 \in \left[\frac{1}{4^\kappa};4^\kappa\right] \\ \sqrt{1-\left(\frac{1}{4}\right)^\kappa} & if \quad \varkappa^2 < \frac{1}{4^\kappa} \end{cases} \tag{99}$$

□

C. *Proof of Proposition 3*

We add two groups of investors with the same information processing capacity to the model of section II.A. Group one with relative magnitude λ and optimal information allocation ρ_1^* and group two with relative magnitude $1-\lambda$ and optimal information allocation ρ_2^*. The resulting optimal asset allocations for each agent group are:

$$q_1^* = \frac{(1-\beta)}{\gamma}A^{-1}\left(\rho_1^{*2}\cdot(s_{1t}-\mu)+\frac{\mu}{1-\beta}-p_t\right) \tag{100}$$

$$q_2^* = \frac{(1-\beta)}{\gamma}A^{-1}\left(\rho_2^{*2}\cdot(s_{2t}-\mu)+\frac{\mu}{1-\beta}-p_t\right) \tag{101}$$

In equilibrium the following equation has to hold, since all assets, normalised to one, have to be held by the agents:

$$1 = \frac{(1-\beta)}{\gamma}A^{-1} \left(\lambda\rho_1^{*2}\cdot(s_{1t}-\mu)+(1-\lambda)\,\rho_2^{*2}\cdot(s_{2t}-\mu)+\frac{\mu}{1-\beta}-p_t\right) \tag{102}$$

Rearranging terms leads to the following price formula:

$$p_t = \lambda\rho_1^{*2}\cdot(s_{1t}-\mu)+(1-\lambda)\,\rho_2^{*2}\cdot(s_{2t}-\mu) + \frac{\mu}{1-\beta}-\frac{\gamma}{(1-\beta)}A1 \tag{103}$$

$$= \rho^{*2}\left((\lambda s_{1t}+(1-\lambda)\,s_{2t})-\mu\right)+\frac{\mu}{1-\beta}-\frac{\gamma}{(1-\beta)}A1 \tag{104}$$

The variance A is given by:

$$A = \sigma^2(1-\rho^{*2}) + \beta\left(\sigma^2 + \left(\lambda^2 + (1-\lambda)^2\right)\tilde{\sigma}^2\right)\rho^{*4} \tag{105}$$

This collapses to the matrix of the non heterogeneous case for $\lambda = 0$ and $\lambda = 1$. Note that $\rho_1^* = \rho_2^* = \rho^*$.

Generalizing to G equally large groups with independent signals one obtains the price as:

$$p_t = \rho^{*2}\left(\sum_{n=1}^{G}\frac{1}{G}s_{nt} - \mu\right) + \frac{\mu}{1-\beta} - \frac{\gamma}{2(1-\beta)}A1 \tag{106}$$

The variance A with G groups is given by:

$$A = \sigma^2(1-\rho^{*2}) + \beta\left(\sigma^2 + \frac{1}{G}\tilde{\sigma}^2\right)\rho^{*4} \tag{107}$$

Thus the price variance for asset one is given by:

$$Var\left(p_t\right) = \left(\sigma^2 + \frac{1}{G}\tilde{\sigma}^2\right)\rho^{*4} \tag{108}$$

□

The variance of the price is therefore not only a question of attention allocation but also a question of how many independent opinions are present. Higher volatility in distress situation may not only result from more attention allocation but also because of more homogeneity, meaning less independent groups.

D. *Proof of Proposition 4*

Within our framework the household problem for the individual agent is given by:

$$\max_{s_t^i\in\Gamma}\mathbb{E}\left[u(c_t^i;s_t^i) + \beta u(c_{t+1}^i;s_t^i)\middle|s_t^i\right] \tag{109}$$

subject to the following constraints

$$\mathbb{I}\left(\{d_t\};\{s_t^i\}\right) \leq \kappa^i \tag{110}$$

$$q_{t+1}^{i*} = \arg\max_{q_{t+1}^i}\mathbb{E}\left[u(c_t^i;s_t^i) + \beta u(c_{t+1}^i;s_t^i)\middle|s_t^i\right] \tag{111}$$

$$c_t^i = q_{t+1}^{i*}\left(d_t - p_t\right) \tag{112}$$

$$c_{t+1}^i = q_{t+1}^{i*}p_{t+1} \tag{113}$$

The agent has mean-variance utility:

$$u(c) = \mathbb{E}[c] - \frac{\gamma}{2} Var[c] \tag{114}$$

Solving this household problem analogous to Proposition 2 yields the following FOC for the quantity of assets the agent wants to hold:

$$\begin{aligned}
0 &= \mathbb{E}[d_t|s_t] + \beta\mathbb{E}[p_{t+1}|s_t] - p_t - \gamma A q^* && (115)\\
0 &= \rho_1^{*2}(s_{1t} - \mu) + \rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t && (116)\\
&\quad + \beta\mathbb{E}[p_{t+1}|\rho] - p_t - \gamma A q^* \\
0 &= \rho_1^{2}(s_{1t} - \mu) + \rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t && (117)\\
&\quad + \sum_{i=1}^{N} \beta^i \left(\rho_2^{*2}(s_{2t} - \mu_t) + \mu - \gamma A q^*\right) \\
&\quad + \beta^{N+1}(\mathbb{E}[p_{t+N+2}]\gamma A q^*) - p_t - \gamma A q^* \\
0 &= \rho_1^{*2}(s_{1t} - \mu) + \sum_{i=0}^{N} \beta^i \left(\rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t - \gamma A q^*\right) && (118)\\
&\quad - p_t
\end{aligned}$$

We assume updating of the agent's best guess μ_t about $\tilde{\mu}$. The agent forms his new opinion on $\tilde{\mu}$ by weighting his signal on the mean s_{2t} and his previous belief μ_{t-1} by the quality of the signal he receives, thus $\mu_t = \rho_2^2 s_{2t} + (1-\rho_2^2)\mu_{t-1}$. The variance of $\tilde{\mu}$ is assumed to be σ_2^2 for each generation to ensure a stationary problem. This results in the optimal asset allocation:

$$\begin{aligned}
q^* &= \frac{(1-\beta)}{\gamma} A^{-1} && (119)\\
&\quad \left(\rho_1^{*2}(s_{1t} - \mu) + \frac{\rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t}{1-\beta} - p_t\right)
\end{aligned}$$

In equilibrium the agent group has to hold all assets, normalized to one:

$$\begin{aligned}
1 &= \frac{(1-\beta)}{\gamma} A^{-1} && (120)\\
&\quad \left(\rho_1^{*2}(s_{1t} - \mu) + \frac{\rho_2^{*2}(s_{2t} - \mu_t) + \mu + \mu_t}{1-\beta} - p_t\right)
\end{aligned}$$

Solving for the price leads to:

$$p_t = \rho_1^{*2}(s_1 - \mu) + \frac{\rho_2^{*2}(s_2 - \mu_t) + \mu + \mu_t}{1-\beta} - \frac{\gamma}{(1-\beta)} A1 \tag{121}$$

The variance A is given by:

$$A = \sigma_1^2(1-\rho_1^{*2}) + \frac{\sigma_2^2}{(1-\beta)^2}(1-\rho_2^{*2}) + \beta\sigma_1^2\rho_1^{*2} + \frac{\beta}{(1-\beta)^2}\rho_2^{*2}\sigma_2^2 \tag{122}$$

With a closed form solution for the price at hand we can turn to the attention allocation problem. Analogue to the proof of proposition 2 the simplified optimization problem ignoring all constant parameters is given by:

$$\max_{\{\rho_1,\rho_2\}\in[0,1]} \sigma_1^2\rho_1^2 + \frac{\sigma_2^2}{(1-\beta)^2}\rho_2^2 \tag{123}$$

This is equivalent to the step 4 in the proof of proposition 2 when simply replacing σ_2 by $\frac{\sigma_2}{(1-\beta)}$. □

E. Proof of Proposition 6

In equilibrium expected prices for all future periods are identical:

$$\tilde{\mu} = \mu_{t+i}\forall i \in \mathbb{N} \rightarrow \mathbb{E}\left[R_t \middle| \tilde{\mu} = \mu_t\right] - (\tilde{\mu} + \mu_t) = 0 \tag{124}$$

During a (positive) shift, which implies non perfect information processing meaning $\rho_2^2 < 1$ and $\tilde{\mu} > \mu_t, \mu_{t+1}$, the expected excess return is given by:

$$\begin{aligned} \mathbb{E}\left[R_t \middle| shift\right] - (\tilde{\mu} + \mu_t) &= \mathbb{E}\left[\frac{\rho_2^2(s_{2t+1} - \mu_{t+1}) + \mu + \mu_{t+1}}{1-\beta} - \frac{\gamma}{(1-\beta)}A1 - \frac{\rho_2^2(s_{2t} - \mu_t) + \mu + \mu_t}{1-\beta} + \frac{\gamma}{(1-\beta)}A1\right] & (125) \\ &= \frac{1}{1-\beta}\rho_2^2(\tilde{\mu} - \mu_{t+1}) & (126) \end{aligned}$$

Since $\tilde{\mu} > \mu_{t+1}$ holds because of the ongoing shift equation (126) will be positive during a (positive) shift. □

During a negative shift the same arguments hold but it implies a negative expected excess return. The model does not imply a possible excess return for the agent on an individual asset, since he does not have the shift information (which is actually forward looking) in his signal, which encompasses all the information he can acquire.

F. N Asset Allocation

$$\max_{\xi_n \in [0,1]} \sum_{n=1}^{N} \sigma_n^2 \xi_n \tag{127}$$

$$\sum_{n=1}^{N} \frac{1}{2} \log_2 \left(\frac{1}{1-\xi_n} \right) \leq \kappa \tag{128}$$

We first look at the general case of a non-corner solution. Simplifying the boundary condition by only taking into account first order and first order interaction effects we obtain:

$$\sum_{n=1}^{N} \xi_n - \sum_{n=1}^{N} \sum_{j=n+1}^{N} \xi_n \xi_j \quad = \quad \alpha \tag{129}$$

As before α is given by:

$$\alpha \quad = \quad 1 - \frac{1}{\exp\left(2\kappa \ln(2)\right)} \tag{130}$$

In case of two assets this is equal to:

$$\xi_1 + \xi_2 - \xi_1 \xi_2 - \alpha = 0 \tag{131}$$

The FOC for a maximum, meaning setting the normal of the differentiable manifold equal to a multiple of the gradient of the target function, is given by:

$$\begin{pmatrix} 0 & 1 & ... & 1 & \sigma_1^2 \\ 1 & 0 & ... & 1 & \sigma_2^2 \\ & & ... & & \\ 1 & ... & 1 & 0 & \sigma_{N^2} \end{pmatrix} \begin{pmatrix} \xi_1 \\ ... \\ \xi_N \\ \lambda \end{pmatrix} = \begin{pmatrix} 1 \\ ... \\ ... \\ 1 \end{pmatrix} \tag{132}$$

We define:

$$\Psi = \begin{pmatrix} 0 & 1 & ... & 1 \\ 1 & 0 & ... & 1 \\ ... & ... & ... & ... \\ 1 & ... & 0 & 1 \\ 1 & ... & 1 & 0 \end{pmatrix} \tag{133}$$

Thus we obtain:

$$\xi(\lambda) = \Psi^{-1}\begin{pmatrix}1\\ \cdots\\ \cdots\\ 1\end{pmatrix} - \lambda\begin{pmatrix}\sigma_1^2\\ \cdots\\ \cdots\\ \sigma_N^2\end{pmatrix} \tag{134}$$

λ^* is given by the solution to the following quadratic equation:

$$\sum_{i=1}^{N}\xi_i(\lambda) - \sum_{i=1}^{N}\sum_{j=i+1}^{N}\xi_i(\lambda)\xi_j(\lambda) = \alpha \tag{135}$$

We obtain the optimal allocation as:

$$\xi^*(\lambda^*) = \Psi^{-1}\begin{pmatrix}1\\ \cdots\\ \cdots\\ 1\end{pmatrix} - \lambda^*\begin{pmatrix}\sigma_1^2\\ \cdots\\ \cdots\\ \sigma_N^2\end{pmatrix} \tag{136}$$

□

For $N = 1$ this trivially leads to $\xi^* = \alpha$ and for $N = 2$ in case of an inner solution to:

$$\lambda^* = \frac{\sqrt{1-\alpha}}{\sigma_1\sigma_2} \tag{137}$$

$$\xi^* = \begin{pmatrix}1-\sqrt{1-\alpha}\frac{\sigma_2}{\sigma_1}\\ 1-\sqrt{1-\alpha}\frac{\sigma_1}{\sigma_2}\end{pmatrix} \tag{138}$$

This is identical to the model without approximation for the one and the two asset case, since the approximation is exact for up to two assets because only for three and more assets higher order interaction terms exist, which are lost due to the approximation.

G. Proof heterogenous information processing capacities

The derivation to proposition (7) of the equilibrium in the case of heterogeneity among agents if they have different information processing capacities follows the same steps as in the baseline model (see proof of proposition 2):

Given any optimal attention allocation the optimal asset allocation of both investor groups is given by:[26]

$$q_1^* = \frac{(1-\beta)}{\gamma}\left(A_1 + \tilde{A}\right)^{-1}\left(\Xi_1^2 \cdot (s_{1t} - \mu) + \frac{\mu}{1-\beta} - p_t\right) \tag{139}$$

$$q_2^* = \frac{(1-\beta)}{\gamma}\left(A_2 + \tilde{A}\right)^{-1}\left(\Xi_2^2 \cdot (s_{2t} - \mu) + \frac{\mu}{1-\beta} - p_t\right) \tag{140}$$

A_1 and A_2 are the variances associated with the dividends for each of the investor groups and $\tilde{A}$ the variances of future prices. In order to shorten notation we define τ_i as:

$$\tau_1 = \frac{(1-\beta)}{\gamma}\left(A_1 + \tilde{A}\right)^{-1} \tag{141}$$

$$\tau_2 = \frac{(1-\beta)}{\gamma}\left(A_2 + \tilde{A}\right)^{-1} \tag{142}$$

In equilibrium the sum of all assets has to equal the total supply of one:

$$1 = \lambda\tau_1\left(\Xi_1^2 \cdot (s_{1t} - \mu) + \frac{\mu}{1-\beta} - p_t\right) \tag{143}$$

$$+ (1-\lambda)\,\tau_2\left(\Xi_2^2 \cdot (s_{2t} - \mu) + \frac{\mu}{1-\beta} - p_t\right) \tag{144}$$

Solving for the price this leads to:

$$p_t = \frac{\mu}{1-\beta} + \omega\left(\Xi_1^2 \cdot (s_{1t} - \mu)\right) + (1-\omega)\left(\Xi_2^2 \cdot (s_{2t} - \mu)\right) - \Omega 1 \tag{145}$$

with

$$\omega = \left(\lambda\tau_1 + (1-\lambda)\,\tau_2\right)^{-1}\lambda\tau_1 \tag{146}$$

$$\Omega = \left(\lambda\tau_1 + (1-\lambda)\,\tau_2\right)^{-1} \tag{147}$$

[26]For the attention allocation process a fixed trading strategy $\bar{q} = 1$ needs to be assumed as a technical assumption.

Thus the variance matrices for the two investor groups are given by:

$$\tilde{A} = \beta[\left(\omega\Xi_1^2 + (1-\omega)\,\Xi_2^2\right)^2 \tag{148}$$

$$+ \left(\omega^2\left(\Xi_1^2 - \Xi_1^4\right) + (1-\omega)^2\left(\Xi_2^2 - \Xi_2^4\right)\right)] \begin{pmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{pmatrix} \tag{149}$$

$$A_i = \begin{pmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{pmatrix} \left(1 - \begin{pmatrix} \rho_{i1}^{*2} & \\ & \rho_{i2}^{*2} \end{pmatrix}\right) \tag{150}$$

The values for ω are implicitly given as the solution to the following equation:

$$\omega = \left(\lambda\left(A_1 + \tilde{A}\right)^{-1} + (1-\lambda)\left(A_2 + \tilde{A}\right)^{-1}\right)^{-1} \lambda\left(A_1 + \tilde{A}\right)^{-1} \tag{151}$$

Unfortunately there exists no closed form solution to equation (151) and thus the value of ω has to be obtained numerically. Since however ω can be pre-computed the loss as compared to a closed form solution is rather small.

H. Proof shift in the long-term

This is the proof to proposition (8). In addition to proposition (7), each group of agents has now current beliefs μ_{1t}, μ_{2t} of the long-term mean, which they update every period. This allows for heterogeneity in the dynamics in the long-term mean expectations. The optimal asset allocation of both groups is given by:

$$q_1^* = \frac{(1-\beta)}{\gamma}\left(A_1+\tilde{A}\right)^{-1} \left(\rho_1^{*2}(s_{11t}-\mu)+\frac{\rho_2^{*2}(s_{12t}-\mu_t)+\mu+\mu_{1t}}{1-\beta}-p_t\right) \tag{152}$$

$$q_2^* = \frac{(1-\beta)}{\gamma}\left(A_2+\tilde{A}\right)^{-1} \left(\rho_1^{*2}(s_{21t}-\mu)+\frac{\rho_2^{*2}(s_{22t}-\mu_t)+\mu+\mu_{2t}}{1-\beta}-p_t\right) \tag{153}$$

Using the same notational short cuts as before the price will be:

$$\begin{aligned} p_t &= \omega\left(\rho_1^{*2}(s_{11t}-\mu)+\frac{\rho_2^{*2}(s_{12t}-\mu_t)+\mu+\mu_{1t}}{1-\beta}\right) \\ &\quad +(1-\omega)\left(\rho_1^{*2}(s_{21t}-\mu)+\frac{\rho_2^{*2}(s_{22t}-\mu_t)+\mu+\mu_{2t}}{1-\beta}\right) \\ &\quad -\Omega_1 \end{aligned} \tag{154}$$

with

$$\omega = (\lambda\tau_1+(1-\lambda)\tau_2)^{-1}\lambda\tau_1 \tag{155}$$

$$\Omega = (\lambda\tau_1+(1-\lambda)\tau_2)^{-1} \tag{156}$$

Thus the variances for the two investor groups are given by:

$$\begin{aligned} \tilde{A} &= \left(\omega\rho_{11}^{*2}+(1-\omega)\rho_{21}^{*2}\right)^2\beta\sigma_1^2 \\ &\quad +\left(\omega^2\frac{\rho_{12}^{*4}}{(1-\beta)^2}+(1-\omega)^2\frac{\rho_{22}^{*4}}{(1-\beta)^2}\right)\beta\sigma_2^2 \\ &\quad +\omega^2\beta\left(\rho_{11}^{*4}\tilde{\sigma}_{11}^2+\frac{\rho_{12}^{*4}}{(1-\beta)^2}\tilde{\sigma}_{12}^2\right) \\ &\quad +\left(1-\omega^2\right)\beta\left(\rho_{21}^{*4}\tilde{\sigma}_{21}^2+\frac{\rho_{22}^{*4}}{(1-\beta)^2}\tilde{\sigma}_{22}^2\right) \end{aligned} \tag{157}$$

$$A_i = \sigma_1^2\left(1-\rho_{i1}^{*2}\right)+\frac{\sigma_2^2}{1-\beta}\left(1-\rho_{i2}^{*2}\right) \tag{158}$$

As with the model without shifts the values for ω are implicitly given as the solution to the following equation:

$$\omega = \left(\lambda\left(A_1+\tilde{A}\right)^{-1}+(1-\lambda)\left(A_2+\tilde{A}\right)^{-1}\right)^{-1}\lambda\left(A_1+\tilde{A}\right)^{-1} \tag{159}$$

REFERENCES

Bacchetta, P. and VanWincoop, E. (2005). A theory of the currency denomination of international trade. *Journal of International Economics*, 67:295–319.

Barberis, N., Huang, M., and Santos, T. (2001). Prospect theory and asset prices. *The Quarterly Journal of Economics*, 116:1–53.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49:307–343.

Biais, B., Bossaerts, P., and Spatt, C. (2010). Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies*, 23:1503–1543.

Brogaard, J. A. (2010). High frequency trading and its impact on market quality. *Working paper.*

Brunnermeier, M. K. (2009). Bubbles: Entry in new palgrave dictionary of economics. *New Palgrave Dictionary of Economics.*

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance*, 66:1461–1499.

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

DeOliveira, H., Denti, T., Mihm, M., and Ozbek, M. K. (2013). Rationally inattentive preferences. *Working paper.*

Dessein, W., Galeotti, A., and Santos, T. (2013). Rational inattention and organizational focus. *CEPR Discussion Paper*, DP9395.

Falkinger, J. (2007). Attention economies. *Journal of Economic Theory*, 133(1):266–294.

Falkinger, J. (2008). Limited attention as a scarce resource in information-rich economies. *The Economic Journal*, 118:1596–1620.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49:283–306.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10:1–21.

Hefti, A. M. (2013). Attention and competition. *Working paper.*

Hefti, A. M. and Heinke, S. (2015). On the economics of superabundant information and scarce attention. *OEconomia*, 5(1):XXX.

Hellwig, C., Kohls, S., and Veldkamp, L. (2012). Information choice technologies. *American Economic Review*, 102:35–40.

Hellwig, C. and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76:223–251.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22:477–498.

Hong, H., Lim, T., and Stein, J. G. (2000). Bad news travels slowly : Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, 55:265–295.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49:1541–1578.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, pages 1429–1445.

Luo, Y. and Young, E. R. (2010). Asset pricing under information-processing constraints. *Economics Letters*, 107:26–29.

Mackowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803.

Mallaby, S. (2010). *More Money Than God*. The Penguin Press., New York, New York.

Matejka, F. and McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review*, 102:24–29.

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., and Preis, T. (2013). Quantifying wikipedia usage patterns before stock market moves. *Scientific Reports*, 3:1801.

Mondria, J., Wu, T., and Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82:85–95.

Pashler, H. E. (1998). *Attention*. Psychology Press.

Peng, L. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, 40:307–329.

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

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. *Scientific Reports*, 3:1684.

Shleifer, A. (2000). *Inefficient markets: An introduction to behavioral finance.* Oxford university press.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50:665–690.

Sims, C. A. (2005). Rational inattention : A research agenda. *Discussion Paper Series 1: Economic Studies*, 34.

Tutino, A. (2013). Rationally inattentive consumption choices. *Review of Economic Dynamics*, 16:421–439.

VanNieuwerburgh, S. and Veldkamp, L. (2008). Information acquisition and under-diversification. *NBER Working Paper*, 13904.

Veldkamp, L. (2011). *Information Choice in Macroeconomics and Finance.* Princeton University Press, Princeton, New Jersey.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics*, 56:100–124.