Dynamic Attention Behavior under Return Predictability*

Daniel Andrei† Michael Hasler‡

January 4, 2016

Abstract

We consider a dynamic problem of asset and attention allocation where an investor jointly chooses how much to consume and invest, but also how much information to acquire about a risky asset whose expected returns are predictable. Higher attention helps the investor to reduce uncertainty about expected returns, but also to speculate on their future path. These two effects reinforce each other when the stock return predictor moves away from its long-term mean, yielding a U-shaped relationship between the return predictor and attention. Higher persistence of the return predictor further magnifies this effect. Surprisingly, attention decreases with stock return volatility.

*We would like to thank Michael Brennan, Peter Christoffersen, and Chayawat Ornthanalai for insightful suggestions and comments. We are also grateful for comments received from conference/seminar participants at ISF 2015, Princeton University, UCLA Anderson, University of Lausanne, and University of Toronto. Financial support from UCLA and the University of Toronto is gratefully acknowledged.

†UCLA, Anderson School of Management, 110 Westwood Plaza, Suite C420, Los Angeles, CA 90095, USA, daniel.andrei@anderson.ucla.edu; www.danielandrei.info

‡University of Toronto, Rotman School of Management, 105 St. George Street, Toronto, ON, M5S 3E6, Canada, Michael.Hasler@rotman.utoronto.ca; www.utoronto.hasler

1 Introduction

The incomplete-information literature is based on the premise that agents operate under partial knowledge of the economy. Typically, one or several state variables driving the economy are assumed to be unobservable and their evolution is governed by stochastic processes, whose parameters are either fixed or must be estimated.

An important incomplete-information problem arises in the dynamic portfolio choice setting (Merton, 1969, 1971). Because ex-ante expected returns are unobservable, investors need to estimate them using available information. In a seminal paper, Gennotte (1986) shows that this joint *filtering* and *portfolio choice* problem can be solved in two separate steps by first estimating expected returns and then choosing the optimal portfolio conditional on these estimated expected returns.[1] It is however unfortunate that the large subsequent literature building on this result neglects the information acquisition problem and makes the implicit assumption that the set of available information used to estimate expected returns is exogenously given. The objective of this paper is to relax this assumption.

Investigating the *joint* problem of information acquisition and portfolio choice is relevant for at least two reasons. First, the staggering amount of resources devoted nowadays to information-related activities suggests a strong demand for financial information among investors. Solving the dynamic information acquisition problem helps us rationalize the observed demand for information. Second, recent studies show that investors' attention to news is time-varying (Da, Engelberg, and Gao, 2011), counter-cyclical (Lustig and Verdelhan, 2012), and linked to investment strategies (Barber and Odean, 2008).[2] As fluctuations in investors' attention can change beliefs in an instant and thus have immediate and real effects on the economy (Andrei and Hasler, 2015), it is essential to understand what variables drive investors' needs to acquire information and how this relates to portfolio strategies.

In our theoretical model, an agent can invest in one risk-free asset and one risky stock with unobservable expected returns. At each point in time, the investor optimally chooses her consumption, portfolio, and quantity of information needed to estimate expected returns in order to maximize her expected lifetime utility of consumption. She is able to control her investment opportunity set by acquiring, at a cost, news coming from various sources: consulting and forecasting reports, financial analysis, historical time series, or simply the Internet. Information acquisition regulates both the learning and the investment decisions of the investor. By acquiring more accurate information, i.e. by paying more *attention to*

[1] See also Detemple (1986) and Dothan and Feldman (1986) for the same result but in different contexts.

[2] See Sicherman, Loewenstein, Seppi, and Utkus (2015) for an empirical study on the dynamics of investors' attention to their portfolios.

news,[3] the investor is able to better estimate expected returns and therefore to increase her expected utility at the expense of decreasing her current wealth. In other words, the investor faces a dynamic trade-off problem of asset and attention allocation.

We assume that expected returns are a linear function of an observable predictive variable, with an *unobservable* predictive coefficient (Xia, 2001).[4] The investor uses all the available historical data to estimate this predictive coefficient and to construct a forecast of future returns. Beyond historical data, the investor can also choose to improve her estimation by collecting news at a cost. Because we assume perfect learning to be infinitely costly, the investor optimally chooses a finite amount of attention. We characterize the optimal amount of attention and its responsiveness to changes in the state variables of the model.

The investor's need to acquire information is driven by two factors. First, the investor dislikes uncertainty and is therefore willing to spend more resources on information acquisition in more uncertain environments. We define this first attention-inducing factor *hedging against uncertainty about expected returns.* Second, because returns are mean-reverting, the investor speculates on their future path (Kim and Omberg, 1996). This speculative behavior prompts the investor to be more attentive in order to build a better outlook. We define this second attention-inducing factor *speculating on the future path of expected returns.*

The predictive variable (which we call hereafter "dividend yield," although any mean-reverting predictive variable generates the same theoretical results) determines to a large extent the optimal amount of attention to news. When the dividend yield moves away from its long-term mean, the impact of the hedging factor and the speculative factor on investor's attention is substantially magnified. First, when the dividend yield is unusually high or low, the uncertainty about expected returns is large, which strengthens the hedging factor. Simultaneously, the investor anticipates a strong reversal to the mean, which strengthens the speculative factor (i.e., "the trend is your friend"). The two effects reinforce each other resulting in a U-shaped relation between the optimal attention and the dividend yield. This prediction of our model is independent on the utility function specification: it holds for any concave time-additive utility function.

We find that attention to news is a decreasing function of the volatility of *realized returns*, which is assumed to be stochastic in our model (Chacko and Viceira, 2005; Liu, 2007). When realized returns volatility increases, the quality of the information environment deteriorates, which prompts the investor to lower its attention to news. Although the prediction of lower attention during periods of high volatility has not been formulated before, the magnitude of

[3]See Sims (2003), Peng and Xiong (2006), Kacperczyk, Nieuwerburgh, and Veldkamp (2009), Mondria (2010), Mondria and Quintana-Domeque (2013) for a similar interpretation of attention.

[4]See also Kandel and Stambaugh (1996), Barberis (2000), and Brandt, Goyal, Santa-Clara, and Stroud (2005) for equivalent assumptions.

this effect is nonetheless weaker than the one implied by the other drivers of attention.[5]

We investigate the sensitivity of our results to changes in the parameters of the dynamics of the dividend yield. More persistence of the dividend yield means more uncertainty (and thus a stronger hedging factor) but also longer trends (and thus a stronger speculative factor), yielding higher attention. Furthermore, a more volatile dividend yield implies more uncertainty and therefore higher attention paid to news. Finally, the U-shaped relation between attention and the dividend yield becomes asymmetric in the presence of a non-zero correlation between the dividend yield and realized returns. When this correlation is negative (Xia, 2001), attention tends to be higher when the dividend yield is high, which typically corresponds to bad economic times.

We examine the role of risk aversion in this information acquisition problem. We show that investors with high risk aversion pay less attention to news than investors with low risk aversion, because a significantly smaller proportion of their wealth is invested in the stock. Therefore, investors with logarithmic utility function turn out to pay high attention to news even though their portfolios are myopic, i.e. their portfolios' hedging demand is equal to zero. A natural implication of the inverse relation between attention and risk aversion is that the certainty equivalent cost of ignoring news is close to zero for investors with high risk aversion, but is substantial for myopic investors.

We conclude with an empirical analysis of the dependence of attention on the state variables of our model. Consistent with the theoretical predictions, investors' attention is a U-shaped function of the dividend yield, a decreasing function of stock-return volatility, and an increasing function of uncertainty. Furthermore, we provide evidence that the relation between risky asset holdings and the state variables is consistent with theory, suggesting that our framework is likely to provide a realistic description of the dynamic investment and information acquisition problem faced by institutional investors.

This paper contributes to a large literature considering portfolio choice problems with stochastic expected returns, stochastic volatility, incomplete information, and uncertainty about predictability (see Wachter (2010) for a recent survey).[6] In a Bayesian framework similar to ours, Xia (2001) investigates the portfolio choice and hedging motives of an investor

[5]Note that in our model there is a distinction between "expected return volatility" and "realized return volatility." The former depends on uncertainty and can be thought of as a forward looking measure, whereas the latter is a contemporaneous measure. As such, our model predicts that investor's attention strongly increases when the expected return volatility is high but slightly decreases when the realized return volatility is high.

[6]For portfolio selection problems with stochastic expected returns see Kim and Omberg (1996), Brandt (1999), Campbell and Viceira (1999), Ait-Sahalia and Brandt (2001), and Wachter (2002), with stochastic volatility see Chacko and Viceira (2005) and Liu (2007), with incomplete information see Gennotte (1986) and Brennan (1998), and with uncertainty about predictability see Kandel and Stambaugh (1996), Barberis (2000), and Brandt et al. (2005).

who faces uncertainty about the predictive power of the dividend yield. Chen, Ju, and Miao (2014) focus on the portfolio choice of an ambiguity averse investor who is uncertain about whether stock returns are iid or predicted by the dividend yield. As in Xia (2001), the optimal share of wealth invested in the stock is non-monotone in the dividend yield. It increases with the dividend yield when the latter is not too large, and decreases with it when the latter is sufficiently large. Our paper contributes to this literature by focusing on when and why investors choose to pay attention to news and therefore learn about the predictability of future returns. That is, our paper highlights the dynamic relation between attention allocation and asset allocation.

Furthermore, our study builds on the literature which considers costly information acquisition. In this literature, Detemple and Kihlstrom (1987) analyze the problem of an investor who invests in production technologies and in an information market. Building on the seminal work of Sims (2003), Veldkamp (2006a,b) show that costly information acquisition helps explain excess co-movement and the simultaneous increases in emerging markets' media coverage and equity prices. Huang and Liu (2007) consider the portfolio choice and information acquisition problem of an investor facing unobservable expected returns. The investor can *initially* choose, at a cost, the accuracy and frequency of news providing information on expected returns. The authors show that costly information prompts the investor to acquire imperfectly accurate news at a low frequency. None of these studies shed light on the dynamic dependence of attention and risky investments on key economic variables such as the dividend yield, uncertainty, and stock-return volatility. As such, our paper represents a first step in understanding the complex role played by investors' attention to news in financial markets.

The rest of the paper proceeds as follows. Section 2 describes the economy and solves the optimal attention and asset allocation problem of the investor. Section 3 presents a numerical illustration of the results, with parameters adopted from the existing literature. Section 4 performs the empirical analysis. Section 5 concludes. The Appendix contains all proofs.

2 The Model

In this section, we describe the economy and the dynamics of the state variables. We solve the investor's learning problem and characterize her optimal attention to news, optimal portfolio choice, and optimal consumption rule.

Consider an economy with an investor with utility function defined by

$$U(c) = \mathbb{E}\left(\int_0^\infty e^{-\delta s} u(c_s) ds\right), \tag{1}$$

where c_t is the consumption at time t, δ is the subjective discount factor, and $u(c)$ is an increasing and concave function of c differentiable on $(0, \infty)$.

The investor continuously trades one instantaneously risk-free asset paying a rate of return r_t and one risky asset (the *stock*) whose value with reinvested dividends follows

$$\frac{dP_t}{P_t} = \mu_t dt + \sqrt{V_t} dB_{P,t}, \tag{2}$$

where μ_t is the instantaneous expected return on the stock and V_t the instantaneous variance of stock returns.

The investor operates under partial knowledge of the economy. Specifically, the expected return μ_t is unobservable, but the investor knows that it is an affine function of an observable state variable y_t. That is, the expected return satisfies

$$\mu_t = \bar{\mu} + \beta_t (y_t - \bar{y}), \tag{3}$$

where β_t is an *unobservable* predictive coefficient (Xia, 2001). The observable predictive variable y_t and the unobservable predictive coefficient β_t evolve according to the following diffusion processes:

$$dy_t = \lambda_y (\bar{y} - y_t) dt + \sigma_y dB_{y,t} \tag{4}$$

$$d\beta_t = \lambda_\beta (\bar{\beta} - \beta_t) dt + \sigma_\beta dB_{\beta,t}, \tag{5}$$

where we assume that $\bar{y}$, λ_y, σ_y, $\bar{\beta}$, λ_β, and σ_β are known constants. The variance of stock returns is observable and follows a square-root process (Heston, 1993):

$$dV_t = \lambda_V (\bar{V} - V_t) dt + \sigma_V \sqrt{V_t} dB_{V,t}, \tag{6}$$

where $\bar{V}$, λ_V and σ_V are known constants.[7]

Given the dynamics of the state variables described above, the investor's problem consists in inferring the predictive coefficient β_t before choosing an optimal portfolio and consumption rule that maximizes the expected lifetime utility of consumption. We turn now to the

[7]Note that the variance V_t is always observable in continuous-time models because it represents the quadratic variation of returns.

inference problem.

2.1 The Inference Process: Active Learning

The investor has the opportunity to actively learn about return predictability, i.e. to collect arbitrarily accurate information about the predictive coefficient β_t. This is achieved by acquiring a news signal with the following dynamics:

$$ds_t = \beta_t dt + \frac{1}{\sqrt{a_t}} dB_{s,t}, \tag{7}$$

where the term driving the instantaneous variance of the signal, $a_t \geq 0$, is optimally controlled by the investor. The larger the value of a_t, the higher the precision of the signal is. Therefore, we refer to the variable a_t as being the investor's *attention to news*.[8]

For ease of exposition, the five Brownian motions $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, $B_{V,t}$, and $B_{s,t}$ are assumed to be independent. Our setup and solution method allows for a general correlation structure among Brownian motions (see Appendix A.1 for the general correlation structure and Table 1 for a more general calibration).

The investor operates in a stochastic environment (Kim and Omberg, 1996; Wachter, 2002; Liu, 2007), but does not observe expected returns (Brennan, 1998; Xia, 2001; Ziegler, 2003). She makes her decisions at time t based on all the information collected up to this point, $\mathcal{F}_t$. This information includes: realized returns defined in (2), changes in the predictive variable defined in (4), changes in the instantaneous variance of stock returns defined in (6), and changes in the signal defined in (7). This last source of information is the focus of our paper. The key feature is that the investor is able to change her information acquisition policy by controlling the magnitude of the noise in the signal (7) *at any point in time*. This results in a control problem with an endogenous information structure.

The evolution of the predictive coefficient, $d\beta_t$, provided in expression (5), is a linear combination of two unobserved factors: $\beta_t dt$ and $dB_{\beta,t}$. The information available to the investor up to time t helps her to get an estimate of β_t. Let us denote by $\widehat{\beta}_t \equiv \mathbb{E}[\beta_t|\mathcal{F}_t]$ the estimated predictive coefficient and its posterior variance by $\nu_t \equiv \mathbb{E}[(\beta_t - \widehat{\beta}_t)^2|\mathcal{F}_t]$. The

[8] Alternatively, the signal could provide information on the unobservable Brownian motion $B_{\beta,t}$ (Scheinkman and Xiong, 2003; Dumas, Kurshev, and Uppal, 2009; Xiong and Yan, 2010). Our current signal specification follows Detemple (1986) and Veronesi (2000) and can be interpreted as a "first-order" signal, whereas the alternative specification would be a "second-order" signal (see Buraschi and Whelan (2012) for a clarifying discussion). Both specifications yield similar results. For a signal providing information on both β_t and $B_{\beta,t}$, see Detemple and Kihlstrom (1987) and Scheinkman and Xiong (2003).

estimated predictive coefficient $\widehat{\beta}_t$ and the posterior variance ν_t are such that

$$\beta_t \sim N(\widehat{\beta}_t, \nu_t), \tag{8}$$

where $N(m, v)$ denotes the Normal distribution with mean m and variance v. Henceforth, we refer to the estimated predictive coefficient $\widehat{\beta}_t$ and the posterior variance ν_t as the *filter* and the *uncertainty*.

The dynamics of the state variables observed by the investor are obtained from standard filtering results (Liptser and Shiryaev, 2001) and are provided in Proposition 1 below.

Proposition 1. *The dynamics of the observed state variables satisfy*

$$\frac{dP_t}{P_t} = \left(\bar{\mu} + \widehat{\beta}_t(y_t - \bar{y})\right) dt + \begin{bmatrix}\sqrt{V_t} & 0 & 0 & 0\end{bmatrix} d\widehat{B}_t^{\perp} \tag{9}$$

$$ds_t = \widehat{\beta}_t dt + \begin{bmatrix}0 & 0 & 0 & \frac{1}{\sqrt{a_t}}\end{bmatrix} d\widehat{B}_t^{\perp} \tag{10}$$

$$dy_t = \lambda_y(\bar{y} - y_t)dt + \begin{bmatrix}0 & \sigma_y & 0 & 0\end{bmatrix} d\widehat{B}_t^{\perp} \tag{11}$$

$$dV_t = \lambda_V(\bar{V} - V_t)dt + \begin{bmatrix}0 & 0 & \sigma_V\sqrt{V_t} & 0\end{bmatrix} d\widehat{B}_t^{\perp}. \tag{12}$$

The dynamics of the filter and of the uncertainty are given by

$$d\widehat{\beta}_t = \lambda_\beta(\bar{\beta} - \widehat{\beta}_t)dt + \begin{bmatrix}\frac{\nu_t(y_t - \bar{y})}{\sqrt{V_t}} & 0 & 0 & \nu_t\sqrt{a_t}\end{bmatrix} d\widehat{B}_t^{\perp} \tag{13}$$

$$\frac{d\nu_t}{dt} = -\left(\frac{(y_t - \bar{y})^2}{V_t} + a_t\right)\nu_t^2 - 2\lambda_\beta\nu_t + \sigma_\beta^2, \tag{14}$$

where $\widehat{B}_t^{\perp} \equiv \begin{bmatrix}\widehat{B}_{1,t}^{\perp} & \widehat{B}_{2,t}^{\perp} & \widehat{B}_{3,t}^{\perp} & \widehat{B}_{4,t}^{\perp}\end{bmatrix}^\top$ *is a 4-dimensional vector consisting of independent Brownian motions observable under investor's filtration.*

Proof. See Appendix A.1. □

Equations (13) and (14) describe investor's updating rule regarding the expectation and variance of the predictive coefficient. The instantaneous change in the filter is driven by four sources of information: realized returns, changes in the predictive variable, changes in volatility, and changes in the news signal. As Equation (13) shows, the investor assigns stochastic weights to these four sources of information.[9] As we will describe below, the size of the weights depend on the relative informativeness of each sources of information.

[9]Because the Brownian motions $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, $B_{V,t}$, and $B_{s,t}$ are uncorrelated, shocks to the predictive variable y_t and to return variance V_t do not impact the investor's estimate of β_t. Hence, the second and third components of the diffusion of $\widehat{\beta}_t$ are both equal to zero. In the general model described in Appendix A.1, and also in our numerical calibration, all four diffusion components are non-zero because the Brownian motions are correlated.

The informativeness of the signal depends on investor's attention, which impacts learning in two ways. First, it has a direct impact on the instantaneous volatility of the filter. Second, it drives the drift of uncertainty. We analyze each of these two effects separately. To facilitate our discussion, we refer to $d\widehat{B}^{\perp}_{1,t}$ as *return shocks* and to $d\widehat{B}^{\perp}_{4,t}$ as *news shocks*.

2.1.1 The impact of attention on the filter

The magnitude of the impact of return shocks and news shocks on the filter depend on the uncertainty ν_t, on the difference between the predictive variable and its long-term mean $y_t - \bar{y}$, and on investor's attention a_t. The following example provides intuition on how the investor updates her beliefs using each piece of information.

Suppose that $y_t > \bar{y}$. Then, an unexpectedly high return ($d\widehat{B}^{\perp}_{1,t} > 0$) means that the current estimate of β_t is too low, and the investor adjusts $\widehat{\beta}_t$ upwards. The opposite happens when $y_t < \bar{y}$: an unexpectedly high return means that the current estimate of β_t is too high, and the investor adjusts $\widehat{\beta}_t$ downwards. Hence, the first coefficient in the diffusion of the filter has the same sign as $y_t - \bar{y}$ (see also Xia (2001) for a similar interpretation).

An additional component drives the filter through active learning from news shocks. When attention is high, the signal becomes more informative and therefore the investor increases the weight assigned to news shocks, as can be seen from the last coefficient in the diffusion of the filter. The instantaneous variance of the filter is an increasing function of attention:

$$\mathrm{Var}[d\widehat{\beta}_t] = \nu_t^2 \left(\frac{(y_t - \bar{y})^2}{V_t} + a_t \right). \tag{15}$$

As attention converges to infinity, the variance of the filter reaches its upper bound σ_β^2. This upper bound represents the variance of the filter when the predictive coefficient β_t is perfectly observable.

2.1.2 The impact of attention on uncertainty

Equation (14) describes the change in uncertainty when the investor controls her attention to news. Uncertainty is locally deterministic and decreases faster when investor's attention is high. The decline in uncertainty is weaker when the predictive coefficient is more persistent (i.e. low λ_β) or when the volatility of realized returns V_t is high. Finally, the last term in (14) shows that the larger the volatility of the predictive coefficient, the stronger the increase in uncertainty over time.

The predictive variable y_t is a key driver of the dynamics of uncertainty. Intuitively, if y_t is close to its long-term mean, learning from realized returns becomes ineffective in estimating β_t because the signal-to-noise ratio is very low. Therefore, the reduction in uncertainty is weak when $y_t - \bar{y} \approx 0$. In contrast, when y_t is far away from its long-term mean, realized returns offer valuable information on the predictive coefficient β_t and thus uncertainty decreases faster.

It is worth noting that uncertainty does not converge to a "steady state" in this model because three stochastic variables, namely the predictive variable y_t, the volatility of asset returns V_t, and investor's attention a_t, drive its dynamics.

2.2 Properties of the estimated expected returns

The following lemma describes the properties of the *estimated expected return* of the stock, defined as:

$$\widehat{\mu}_t = \bar{\mu} + \widehat{\beta}_t(y_t - \bar{y}). \tag{16}$$

Lemma 1. *The dynamics of the estimated expected return follow*

$$d\widehat{\mu}_t = (\lambda_y + \lambda_\beta)\left(\bar{\mu} + \frac{\bar{\beta}\lambda_\beta(y_t - \bar{y})}{\lambda_y + \lambda_\beta} - \widehat{\mu}_t\right)dt + \left[\tfrac{(y_t-\bar{y})^2\nu_t}{\sqrt{V_t}} \quad \sigma_y\widehat{\beta}_t \quad 0 \quad \nu_t\sqrt{a_t}(y_t - \bar{y})\right]d\widehat{B}_t^{\perp}. \tag{17}$$

The mean square error of this estimate (i.e. the uncertainty about expected returns, which we denote hereafter by η) satisfies

$$\eta_t \equiv \mathbb{E}\left[(\mu_t - \widehat{\mu}_t)^2 \middle| \mathcal{F}_t\right] = (y_t - \bar{y})^2\nu_t. \tag{18}$$

The instantaneous variance of the estimated expected return is

$$\mathrm{Var}[d\widehat{\mu}_t] = \nu_t^2(y_t - \bar{y})^2\left(a_t + \frac{(y_t - \bar{y})^2}{V_t}\right) + \sigma_y^2\widehat{\beta}_t^2. \tag{19}$$

$\mathrm{Var}[d\widehat{\mu}_t]$ *is a monotone increasing function of attention. Its maximum depends on both* $\widehat{\beta}_t$ *and* y_t *and is given by*

$$\lim_{a_t \to \infty} \mathrm{Var}[d\widehat{\mu}_t] = \sigma_\beta^2(y_t - \bar{y})^2 + \sigma_y^2\widehat{\beta}_t^2. \tag{20}$$

Expected returns mean-revert at speed $\lambda_y + \lambda_\beta$ to a stochastic level that depends on the predictive variable. If the long-term mean $\bar{\beta}$ is assumed to be zero—which means that there is no predictability on average—then the stochastic level simplifies to the constant $\bar{\mu}$.

As shown in Equation (17), when the filter $\widehat{\beta}_t$ is large, expected returns react strongly to changes in the predictive variable y_t (second component of the diffusion). Furthermore, if investor's attention is high, expected returns react strongly to news and to return shocks, but *only* when $y_t \neq \bar{y}$ (first and fourth component of the diffusion). This concurs with the relation between returns and the predictive variable described in (3), whereby more news on the predictive coefficient β_t—no matter how accurate—are not going to change investor's view about expected returns if $y_t - \bar{y} = 0$.

When $y_t - \bar{y} \neq 0$, an increase in uncertainty increases the variance of expected returns, which now respond more aggressively to both return and news shocks. This is shown in Equation (19). Furthermore, higher attention (or more accurate news) increases the variance of expected returns by making them more sensitive to news shocks. The variance of expected returns reaches its maximum when attention converges to infinity, as shown in Equation (20). This latter equation is obtained by applying Itô's lemma to the expected return in (3) and by assuming that the predictive coefficient β_t is perfectly observable.

To summarize, attention drives two important factors, the variance of expected returns and the drift of uncertainty. More attention increases the sensitivity of expected returns to news shocks and therefore augments their variance. In the same time, more attention yields lower future uncertainty by magnifying the negative component of its drift.

2.3 Optimal Attention, Portfolio Choice, and Consumption

Turning now to investor's optimization problem, we consider $\widehat{\mu}_t$ as a state variable instead of $\widehat{\beta}_t$, with the aim to better interpret and characterize our results. Given this, the investor's problem is to choose consumption c_t, attention to news a_t, and the risky investment share w_t so as to maximize her expected lifetime utility of consumption conditional on her information set at time t, $\mathcal{F}_t$. That is, the investor's maximization problem writes

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) \equiv \max_{c,a,w} \mathbb{E}\left[\int_t^\infty e^{-\delta(s-t)} u(c_s) ds \middle| \mathcal{F}_t\right], \tag{21}$$

subject to the budget constraint

$$dW_t = \left[(r_t - K(a_t))W_t + w_t W_t\left(\widehat{\mu}_t - r_t\right) - c_t\right] dt + w_t W_t \begin{bmatrix} \sqrt{V_t} & 0 & 0 & 0 \end{bmatrix} d\widehat{B}_t^{\perp}. \tag{22}$$

We assume that the total information cost, $W_t K(a_t)$, is linear in wealth. This assumption preserves the homogeneity of the value function and simplifies considerably the problem.[10]

[10] Liu, Peleg, and Subrahmanyam (2010) show that the more wealthy the investor is, the more valuable (in

The per-unit of wealth cost function, $K(a_t)$, is assumed to be increasing and convex in attention. This implies that perfect information ($a_t \to \infty$) cannot be attained, and thus the investor is never able to observe the true level of expected returns. If the investor chooses to pay no attention to news ($a_t = 0$), and therefore to learn using information provided by the price P_t, the predictive variable y_t, and the variance of stock returns V_t only, then her entire wealth is invested in the financial market. In contrast, if the investor decides to pay attention to news ($a_t > 0$), then a positive fraction of her wealth flows to the information market in order to pay for research expenditures. Attention, therefore, can be perceived as a non-financial security in the investor's portfolio.

The dynamics of the vector of state variables $Z_t \equiv [\widehat{\mu}_t, y_t, V_t, \nu_t]^\top$ satisfy

$$dZ_t = m_t dt + \Sigma_{1,t} d\widehat{B}_{1,t}^{\perp} + \Sigma_{2,t} \begin{pmatrix} d\widehat{B}_{2,t}^{\perp} \\ d\widehat{B}_{3,t}^{\perp} \\ d\widehat{B}_{4,t}^{\perp} \end{pmatrix}, \tag{23}$$

where m_t, $\Sigma_{1,t}$, and $\Sigma_{2,t}$ are defined in Appendix A.3. Using this notation, we can write the optimality condition for (21) as

$$0 = -\delta J + \max_{c,a,w} \left(u(c_t) + \mathcal{D}^{W,Z} J \right), \tag{24}$$

where $\mathcal{D}^{W,Z}$ is the infinitesimal generator such that[11]

$$\begin{aligned} \mathcal{D}^{W,Z} J =& J_Z' m + J_W \left[(r_t - K(a_t)) W_t + w_t W_t \left(\widehat{\mu}_t - r_t \right) - c_t \right] && (25) \\ &+ \frac{1}{2} J_{WW} W_t^2 w_t^2 V_t + \frac{1}{2} \mathrm{tr} \left[(\Sigma_1 \Sigma_1' + \Sigma_2 \Sigma_2') J_{ZZ} \right] && (26) \\ &+ W_t w_t \sqrt{V_t} \Sigma_1' J_{WZ}. && (27) \end{aligned}$$

Differentiating (21) partially with respect to the control variables yields the first order conditions:

$$0 = u_c - J_W \tag{28}$$

$$0 = J_W W_t (\widehat{\mu}_t - r_t) + J_{WW} W_t^2 V_t w_t + J_{W\mu} W_t \nu_t (y_t - \bar{y})^2 \tag{29}$$

$$0 = -K'(a_t) J_W W_t - \nu_t^2 J_\nu + \frac{1}{2} \nu_t^2 (y_t - \bar{y})^2 J_{\mu\mu}. \tag{30}$$

dollar terms) information becomes. The reason is that signals typically provide information on returns and not on growth in dollar terms. Our specification of the information cost implies that valuable information rewards equivalently (in net of information costs return terms) wealthy and unwealthy investors, which we believe is perfectly realistic.

[11] Note that, for notational convenience, we drop hats when state variables appear as indices.

Solving the first order conditions yields the optimal consumption c_t^*, risky investment share w_t^*, and attention to news a_t^*:

$$c_t^* = u_c^{-1}(J_W) \tag{31}$$

$$w_t^* = \frac{\widehat{\mu}_t - r_t}{V_t}\frac{-J_W}{J_{WW}W_t} + \frac{\nu_t(y_t - \bar{y})^2}{V_t}\frac{-J_{W\mu}}{J_{WW}W_t} \tag{32}$$

$$a_t^* = \Phi\left(\frac{1}{2J_W W_t}\left(\nu_t^2(y_t-\bar{y})^2 J_{\mu\mu} - 2\nu_t^2 J_\nu\right)\right), \tag{33}$$

where $\Phi(\cdot) \equiv K'(\cdot)^{-1}$ is the inverse of the derivative of the cost function. Because the cost function is increasing and convex in attention, the function $\Phi(\cdot)$ is positive and increasing.

Equation (31) is the standard optimal consumption derived by Merton (1971). The optimal risky investment share, expressed in Equation (32), comprises a myopic and a hedging portfolio (Merton, 1971). The hedging term represents the effect of parameter learning and impacts significantly the asset allocation decision (Brennan, 1998; Xia, 2001). It is positive if $\gamma < 1$ and negative if $\gamma > 1$. It vanishes when the state variables are observable (i.e. when $\nu_t = 0$) because, by assumption, none of these variables are correlated to returns. We discuss the optimal portfolio choice and hedging demand when state variables are correlated to returns in Section 3.2.

Our object of focus is the optimal attention a_t^*, experssed in Equation (33). Since the function $\Phi(\cdot)$ is positive and increasing, we can directly analyze the term in brackets. Attention is driven by two factors: the *uncertainty factor* J_ν, which measures the extent to which the investor dislikes uncertainty, and the *pure state risk aversion factor* $J_{\mu\mu}$, which measures the extent to which the investor (dis)likes variability in estimated expected returns (this factor can be positive or negative, depending on the convexity/concavity of the value function). The uncertainty factor is multiplied by $-\nu_t^2$, which is the marginal effect of attention on the drift of ν_t (see Equation 14). The pure state risk aversion factor is multiplied by $\nu_t^2(y_t - \bar{y})^2$, which is the marginal effect of attention on the variance of expected returns (see Equation 19).

The uncertainty factor J_ν is always negative and implies that the more the investor dislikes uncertainty (i.e. the more negative J_ν), the higher the level of attention aiming at reducing it. In this case, the investor increases her attention (acquires more information) in order to *hedge* against expected returns uncertainty. Furthermore, because our setup features mean-reverting expected returns, the value function is convex in $\widehat{\mu}_t$, which yields $J_{\mu\mu} \geq 0$ (Kim and Omberg, 1996). This induces the investor to take advantage of the convexity and learn more when the return predictor is away from its' long-term mean (i.e., when return predictability is strong). Thus, the more the investor likes convexity (i.e. the more positive

$J_{\mu\mu}$), the higher the level of attention aiming at speculating on it. In this case, the investor increases her attention in order to *speculate* on the future path of expected returns.

Proposition 2 ensures that the optimal attention is always non-negative.

Proposition 2. *Consider two environments that differ only with respect to the amount of attention to news, a_t and a'_t. Assume further that environment a_t is more informative than environment a'_t, i.e. $a_t > a'_t$. Then environment a_t is always more valuable to the investor.*

Proposition 2 is an application of Blackwell's theorem on the equivalence of orderings of experiments. It shows that the investor favors a more informative information structure and has an interest in extracting the maximum amount of information possible given the cost $K(a_t)$. For proofs of this result in various contexts, see Blackwell (1953), Marschak and Miyasawa (1968), Grossman, Kihlstrom, and Mirman (1977), Cremer (1982), and Kihlstrom (1984). In the context of our setup, Proposition 2 implies that

$$\nu_t^2(y_t - \bar{y})^2 J_{\mu\mu} - 2\nu_t^2 J_\nu \geq 0, \tag{34}$$

i.e. investor's attention is always strictly positive or zero.

2.3.1 Reinterpretation in terms of uncertainty about expected returns

According to Lemma 1, the uncertainty about expected returns is defined as $\eta_t \equiv (y_t - \bar{y})^2 \nu_t$. By a change of variable argument, we can therefore rewrite the optimal attention as:[12]

$$a_t^* = \Phi\left(\nu_t^2(y_t - \bar{y})^2 \frac{J_{\mu\mu} - 2J_\eta}{2J_W W_t}\right). \tag{35}$$

The ratio in Equation (35) can be identified as the *need* to acquire more information: it grows with the convexity of J with respect to expected returns and with investor's aversion to uncertainty about expected returns. The need to acquire information is driven by hedging against expected return uncertainty but also by speculating on the future path of expected returns. These two factors are now conveniently grouped into a single term.

The need to acquire information is multiplied by two terms, the uncertainty ν_t and the state variable y_t. These two terms imply that the investor optimally chooses to be more attentive to news when (i) uncertainty is high and (ii) the predictive variable moves away from its long-term mean. As Lemma 1 shows, the marginal benefit of learning about

[12] This functional form might suggest that the optimal attention is zero when $y_t = \bar{y}$, but this is not the case. Indeed, the partial derivative of J with respect to η_t and $\widehat{\mu}_t$ are $J_\eta = J_\nu/(y_t - \bar{y})^2$ and $J_{\mu\mu} = J_{\beta\beta}/(y_t - \bar{y})^2$, respectively. Therefore, simple algebra shows that the "knife-edge" case $y_t = \bar{y}$ does not yield zero attention.

predictability is higher when $|y_t - \bar{y}|$ is large, because significant revisions in expected returns can only be obtained in these states.

Writing the optimal attention under the form (35) is particularly useful because it transparently conveys an amplification mechanism, which is independent on investor's utility function. As expression (35) shows, the effects of uncertainty and the dividend yield on attention reinforce each other, yielding high attention in environments characterized by high uncertainty and by a large difference between the predictive variable and its long-term mean. Overall, this mechanism clearly generates a non-linear relation between uncertainty, the predictive variable, and investor's attention.

2.4 Illustration: CRRA utility and quadratic attention cost

To illustrate the effects of optimal learning about predictability, we assume that the investor has a CRRA utility function with risk aversion parameter γ. In addition, the cost function for attention is specified in quadratic form:

$$K(a_t) = ka_t^2, \tag{36}$$

where $k \geq 0$ is the information cost parameter. In this case, the inverse of the derivative of the cost function satisfies

$$\Phi(x) = \frac{x}{2k}. \tag{37}$$

Under these assumptions, the value function J can be written as follows

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t). \tag{38}$$

Computing the partial derivatives of J as a function of the partial derivatives of ϕ and substituting them into the first-order conditions yields the optimal consumption, risky investment share, and attention to news provided in Proposition 3 below.

Proposition 3. *The optimal consumption c_t^*, risky investment share w_t^*, and attention to* *news a_t^* are given by*

$$c_t^* = \phi^{-1/\gamma} W_t \tag{39}$$

$$w_t^* = \frac{\widehat{\mu}_t - r_t}{\gamma V_t} + \frac{\nu_t (y_t - \bar{y})^2}{\gamma V_t} \frac{\phi_\mu}{\phi} \tag{40}$$

$$a_t^* = \nu_t^2 \left(\frac{\phi_\nu}{\phi} \frac{1}{2k(\gamma - 1)} + \frac{-\phi_{\mu\mu}}{\phi} \frac{(y_t - \bar{y})^2}{4k(\gamma - 1)} \right). \tag{41}$$

The optimal consumption defined in Equation (39) is well-known (Merton, 1971) and does not require any further analysis. The optimal risky investment share defined in Equation (40) is analyzed by Xia (2001) in a setup with constant stock-return volatility.[13] We discuss the dependence of the risky investment share and its hedging components on the state variables in Section 3.2.

The optimal attention defined in Equation (41) is a strictly increasing quadratic function of uncertainty (according to Proposition 2, attention is always non-negative and thus the term multiplying ν_t^2 is strictly positive). We discuss the dependence of investor's attention on the state variables in Section 3.1.

Although an indirect dependence of attention on the variance of stock returns, V_t, arises through the function ϕ, the variance of stock return has no direct impact on investor's attention. In Section 3.1 we show that the indirect impact of the return variance on attention is quantitatively weak, as opposed to the impact of uncertainty and the predictive variable.

3 Numerical Results

In this section, we investigate the properties of optimal attention and optimal risky investment share. We conclude by computing the certainty equivalent cost of not paying attention to news.

We start by calibrating the model. We set the risk aversion coefficient and the subjective discount rate to $\gamma = 3$ and $\delta = 0.01$, respectively. We follow Xia (2001) and consider the dividend yield to be the predictive variable. The dividend yield is defined as $y_t \equiv \log(D_t/P_t)$, where D_t is the dividend paid by the stock at time t, and plays a prominent role in studies of return predictability (Cochrane, 2008; van Binsbergen and Koijen, 2010).[14]

[13] In Xia (2001), the risky investment share features additional terms resulting from the correlations between realized returns and the predictive variable y_t, and between returns and the predictive coefficient β. These correlations are set to zero in the present case. However, our general model (see Appendix A.1) and our calibration feature the same additional terms. Note also that, although a similar decomposition appears in Xia (2001), the attention level affects the shape of the value function, causing differences between portfolio holdings obtained in the two models.

[14] Several other important predictive variables have been identified. They include past market returns,

Parameter	Symbol	Value
Mean-reversion speed of variance	λ_V	6.5
Long-term mean of volatility	$\sqrt{\bar{V}}$	0.15
Volatility of variance	σ_V	0.45
Mean-reversion speed of dividend yield	λ_y	0.2
Long-term mean of dividend yield	$\bar{y}$	0.04
Volatility of dividend yield	σ_y	0.006
Mean-reversion speed of β	λ_β	0.115
Long-term mean of β	$\bar{\beta}$	0
Volatility of β	σ_β	0.3
Long-term expected return	$\bar{\mu}$	0.091
Correlation between β and returns	$\rho_{\beta P}$	0
Correlation between β and dividend yield	$\rho_{\beta y}$	0
Correlation between β and variance	$\rho_{\beta V}$	0
Correlation between returns and dividend yield	ρ_{Py}	-0.93
Correlation between returns and variance	ρ_{PV}	-0.8
Correlation between variance and dividend yield	ρ_{Vy}	0.6

Table 1: Calibration.

In the previous sections, we have set all correlations between the Brownian motions to zero for ease of exposition. To provide a realistic illustration of the relation between attention, the risky investment share, and the state variables, we relax this assumption and consider non-zero correlations (see Appendix A.1 for the full derivation of the model with non-zero correlations). The parameter values are provided in Table 1. Variance parameters are adopted from Christoffersen, Jacobs, and Mimouni (2010), whereas expected return and dividend yield parameters are adopted from Xia (2001). The correlation between returns and return variance is negative (Christoffersen et al., 2010) and reflects the leverage effect (Black, 1976; Christie, 1982). The correlation between returns and the dividend yield is negative (Xia, 2001) because the dividend yield is counter-cyclical. Since the variance of stock returns is also counter-cyclical, we assume a positive correlation between the return variance and the dividend yield. The risk-free rate is constant, $r = 3.4\%$, and the cost parameter is $k = 0.01$.

We solve the partial differential equation resulting from the specification in (38) by applying the Chebyshev collocation method described in Judd (1998). More details on the solution method are provided in Appendix A.2.

the earnings-price ratio, nominal interest rates, and expected inflation. See Goyal and Welch (2008) for a comprehensive survey. See also Ang and Bekaert (2007).

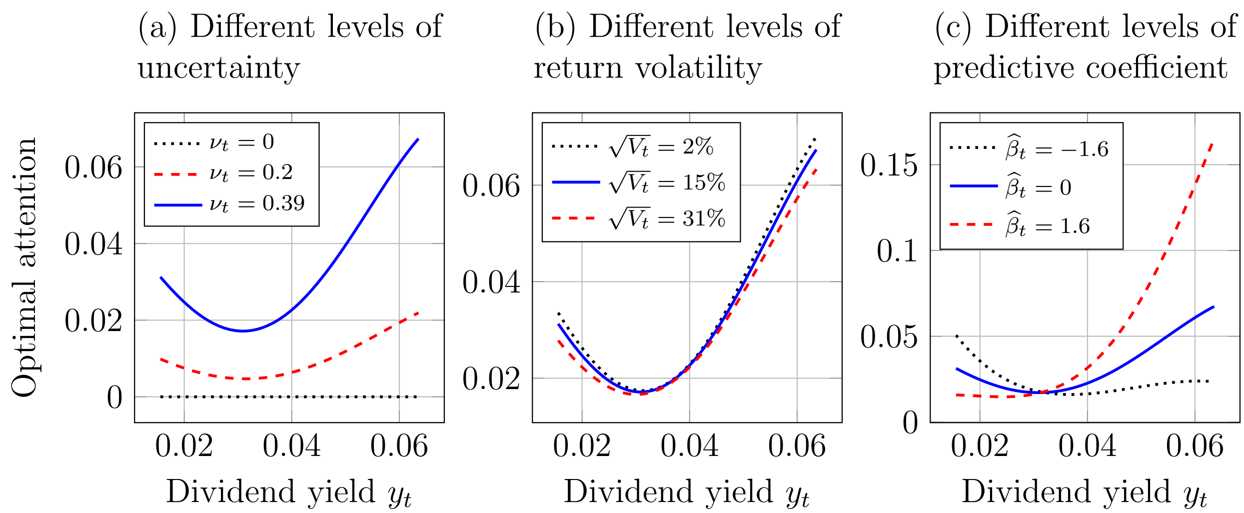

Figure 1: Impact of the dividend yield on attention.

The three panels depict the relation between attention and the dividend yield. We plot three curves corresponding to three different levels of uncertainty in panel (a), return volatility in panel (b), and the predictive coefficient in panel (c). If not stated otherwise, state variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 0.15$, and $\nu_t = \nu_{ss} = 0.39$. Parameter values are provided in Table 1.

3.1 Optimal Attention

Four state variables potentially impact the optimal level of attention: the predictive variable y_t, the uncertainty ν_t, the return variance V_t, and the predictive coefficient $\widehat{\beta}_t$. Since the dividend yield y_t is the main determinant of expected returns, we choose to plot the optimal level of attention as a function of the dividend yield.

Figure 1 reports plots for three different levels of uncertainty in panel (a), return volatility in panel (b), and the estimated predictive coefficient in panel (c). The solid blue line is obtained by setting ν_t, V_t, and $\widehat{\beta}_t$ to their long-term levels ν_{ss}, $\sqrt{\bar{V}}$, and $\bar{\beta}$.[15] The dashed red lines are obtained by setting ν_t to approximately half its long-term level, V_t to its 99.5 percentile, and $\widehat{\beta}_t$ to its 99.5 percentile. The dotted black lines are obtained by setting ν_t to zero, V_t to its 0.5 percentile, and $\widehat{\beta}_t$ to its 0.5 percentile. The solid blue lines are the same in the three panels (although scales are different) and the range of the x-axis in each panel is the 99% confidence interval of the dividend yield.

Figure 1 confirms the result of Proposition 3 and shows that attention is a U-shaped function of the dividend yield. This relationship, which is independent on the utility function

[15]The "steady-state" uncertainty is determined by solving $d\nu_{ss}/\nu_{ss} = 0$ conditional on setting $y_t = \bar{y}$, $V_t = \bar{V}$, and $a_t = 0$ in the dynamics of ν_t, which yields $\nu_{ss} = \sigma_\beta^2/2\lambda_\beta$. This is an upper bound of uncertainty because all the sources of information (i.e. the stock return, the dividend yield, the return variance, and the signal) are uninformative when $y_t = \bar{y}$ and $a_t = 0$.

specification (see expression 35), can be observed uniformly in all three panels.

Panel (a) depicts the optimal attention for three different values of uncertainty. Uncertainty drives investor's attention in two ways: directly through the presence of ν_t^2 in (41) and indirectly through the value function and its derivatives. The higher the uncertainty ν_t, the higher the investor's willingness to spend resources on information acquisition, as can be observed from the graph.

Panel (a) also shows that the impact of uncertainty on attention is further amplified when the dividend yield is far away from its long-term mean ($\bar{y} \neq 0.04$). This implies that high uncertainty coupled with the economy being in states "away-from-normal" endogenously yield high attention to news. In these states, uncertainty about expected returns is high and the investor is more attentive in order to hedge against it. Simultaneously, the investor anticipates a strong reversion to the mean and thus is more attentive in order to speculate on the future path of expected returns. The large difference $|y_t - \bar{y}|$ amplifies both effects and thus generates the U-shape.

Panel (b) depicts a relatively weaker impact of the realized return volatility on attention. According to our discussion in the previous section, return volatility impacts attention only indirectly through the value function, as can also be seen from expression (41). Although the effect is small, panel (a) shows that attention *decreases* when return volatility is high. The investor becomes optimally less attentive when returns are more volatile, because the environment is less informative (i.e., expected returns are less sensitive to new information) and thus the marginal benefit of being more attentive is reduced.[16]

Panel (c) shows the impact of the filter on the relation between attention and the dividend yield. Attention is magnified whenever the product $\widehat{\beta}_t(y_t - \bar{y})$ is positive and large (i.e. when the expected return $\widehat{\mu}_t$ is large). Although $\widehat{\beta}_t$ does not enter directly into expression (41), it has a strong impact through the value function. Two effects arise when the product $\widehat{\beta}_t(y_t - \bar{y})$ is positive and large: first, the investor builds a more aggressive risky investment share; second, expected returns are highly sensitive to new information. These effects reinforce each other and prompt the investor to pay more attention to news when $\widehat{\beta}_t(y_t - \bar{y})$ is positive and large.

The three panels in Figure 1 depict an asymmetric effect: attention tends to be uniformly

[16]This prediction is similar to the "ostrich effect," documented by Galai and Sade (2006), Karlsson, Loewenstein, and Seppi (2009) and Sicherman et al. (2015). The "ostrich effect" states that investors prefer to pay attention to their portfolios following positive news and "put their head in the sand" when they expect to see bad news. Andries and Haddad (2014) provide an alternative rationalization of the "ostrich" effect. In their model, investors are disappointment averse and thus riskier environments encourage more inattention. In our case, riskier returns offer less marginal benefit for being attentive (precisely, it makes the expected returns less responsive to information). Note that the "ostrich effect" commonly refers to *attention to wealth* (Abel, Eberly, and Panageas, 2007, 2013), whereas here we model investors' *attention to news.*

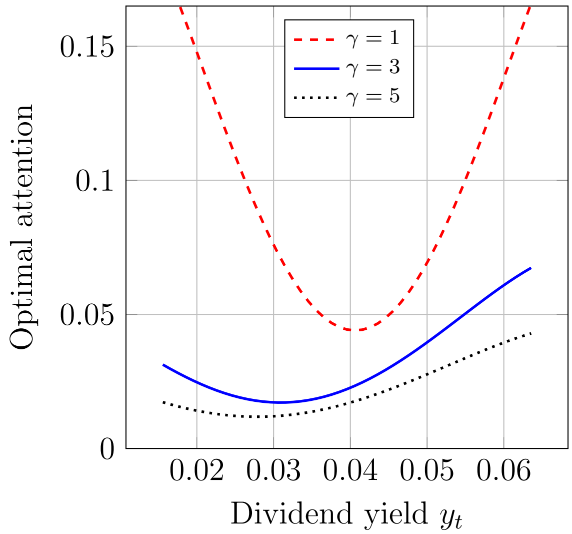

Figure 2: Impact of risk aversion on attention.

The figure depicts the relation between attention and the dividend yield for three different levels of risk aversion. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 0.15$, and $\nu_t = \nu_{ss} = 0.39$. Parameter values are provided in Table 1.

higher when the dividend yield is large. This asymmetry is due to our calibration, which specifies a negative correlation between returns and the dividend yield. This calibration impacts the correlation between realized returns and expected returns and shifts the curves to the left of the long-term dividend yield, generating the asymmetric effect.[17]

We turn now to the relationship between attention and risk aversion. Figure 2 plots the relation between attention and the dividend yield for different levels of risk aversion. The solid blue line in this plot is identical to the solid blue lines in the three panels of Figure 1. An increase in risk aversion (while keeping $\gamma \geq 1$) directly scales down the level of attention, as shown in (41). This first effect is driven by the impact of risk aversion on the optimal portfolio: the larger the risk aversion, the smaller the risky investment share, and therefore the smaller the incentive to gather information on expected returns. A second effect arises through the sensitivities $\phi_\nu > 0$ and $\phi_{\mu\mu} < 0$: a larger risk aversion coefficient amplifies these two sensitivities and prompts the investor to acquire more information.[18] However, the first effect dominates, which results in lower attention levels for more risk-averse investors.

In the log-utility case ($\gamma = 1$), the optimal attention preserves its U-shape, even though the log-utility investor behaves myopically. It is worth mentioning that the asymmetric effect

[17] To see this, fix $\widehat{\beta}_t \neq 0$ and $y_t = \bar{y}$. In this situation, the expected return is $\widehat{\mu} = \bar{\mu}$ and there is no updating from realized return shocks. Suppose now that the dividend yield receives a positive shock (which, by calibration, is negatively correlated with the realized return shock). Following the y_t shock, the investor adjusts the expected return downwards if $\widehat{\beta}_t < 0$ or upwards if $\widehat{\beta}_t > 0$. In any case, expected returns and realized returns become correlated, although there is no learning from realized returns.

[18] Note that $\phi_\nu > 0$ and $\phi_{\mu\mu} < 0$ when $\gamma > 1$ because $J_\nu < 0$ and $J_{\mu\mu} > 0$.

disappears in this case: the attention curve is perfectly centered at $\bar{y} = 0.04$. This shows that the hedging demand is the main channel through which the negative correlation between the dividend yield and realized returns generates an asymmetry in attention.

3.1.1 More on the impact of the predictive variable

We investigate the robustness of our results through a sensitivity analysis of the optimal attention with respect to changes in the dynamics of the predictive variable. More precisely, we analyze how the optimal attention depends on the persistence of the predictive variable, on its volatility, and on its correlation with realized returns.

Figure 3 depicts the optimal attention against the dividend yield for three different levels of persistence of the dividend yield in panel (a), volatility of the dividend yield in panel (b), and correlation between the dividend yield and returns in panel (c). The solid blue lines represent our benchmark calibration (see Table 1) and are the same in the three panels. The dividend yield takes values within its 99% confidence interval. The ranges for the individual lines in panels (a) and (b) are different because the unconditional distribution of the dividend yield depends on its persistence and volatility.

Panel (a) shows that attention increases with persistence and that the U-shaped pattern is more pronounced in more persistent environments. These effects are driven by both the uncertainty factor J_ν and the pure state risk aversion factor $J_{\mu\mu}$. As shown by Kim and Omberg (1996) in a setup with complete information, the mean-reverting property of the risk premium dictates a value function convex in expected returns. This convexity is stronger if expected returns are more persistent.[19] In our setup, a stronger persistence of returns has a double effect. First, it increases the overall uncertainty in the economy—the investor increases her attention in order to hedge this effect. Second, it amplifies the convexity of the value function—the investor increases her attention in order to speculate more on the future path of expected returns.

Panel (b) shows that attention increases with the volatility of the dividend yield. When the dividend yield is more volatile, expected returns are more volatile and the investor pays more attention to news.

Panel (c) confirms that the asymmetric relation between attention and the dividend yield is driven by the correlation between the dividend yield and realized returns. When the correlation approaches zero, shocks to the dividend yield do not induce any "artificial" correlation between realized and expected returns, keeping attention symmetric around $y_t =$

[19] In Kim and Omberg (1996), the value function is convex over a specific interval of expected returns and concave outside the interval. The width of this interval shrinks with the persistence of expected returns, and its bounds are dictated by the two solutions of a quadratic equation.

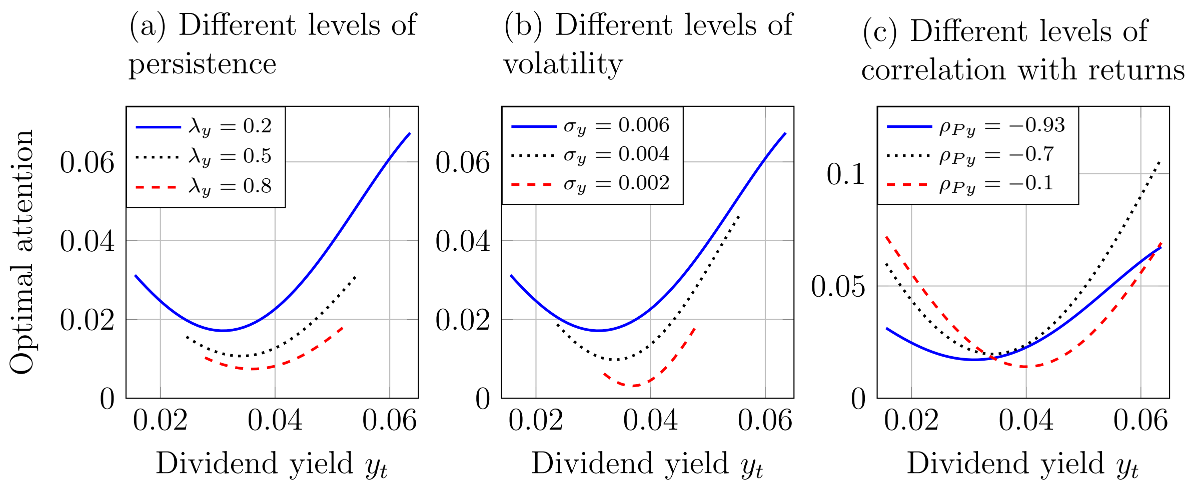

Figure 3: Impact of the dynamics of the dividend yield on attention.
Panels (a), (b), and (c) depict the relation between attention and the dividend yield for three different levels of persistence of the dividend yield, volatility of the dividend yield, and correlation between the dividend yield and returns, respectively. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 0.15$, and $\nu_t = \nu_{ss} = 0.39$. If not stated otherwise, parameter values are provided in Table 1.

$\bar{y}$.

3.2 Optimal Risky Investment Share

Figure 4 plots the optimal risky investment share against the dividend yield for different values of uncertainty in panel (a), return volatility in panel (b), and the estimated predictive coefficient in panel (c). The solid blue line is obtained by setting ν_t, V_t, and $\widehat{\beta}_t$ to their long-term levels ν_{ss}, $\sqrt{\bar{V}}$, and $\bar{\beta}$. The dashed red lines are obtained by setting ν_t to approximately half its long-term level, V_t to its 99.5 percentile, and $\widehat{\beta}_t$ to its 99.5 percentile. The dotted black lines are obtained by setting ν_t to zero, V_t to its 0.5 percentile, and $\widehat{\beta}_t$ to its 0.5 percentile. The solid blue lines are the same in the three panels (although scales are different) and the range of the x-axis in each panel is the 99% confidence interval of the dividend yield.

The risky investment share decreases with both uncertainty (panel a) and return volatility (panel b), because both variables are measures of risk. This is confirmed by Equation (40), which shows that uncertainty loads negatively on the risky investment share, whereas the return variance scales the risky investment share.[20] Furthermore, since the risky investment share increases with expected returns, which depend on the product of the predictive coefficient and the demeaned dividend yield, the risky investment share increases with the

[20]Uncertainty loads negatively on the risky share because $J_\mu > 0$ and $J < 0$ implies that $\phi_\mu < 0$ and $\phi > 0$.

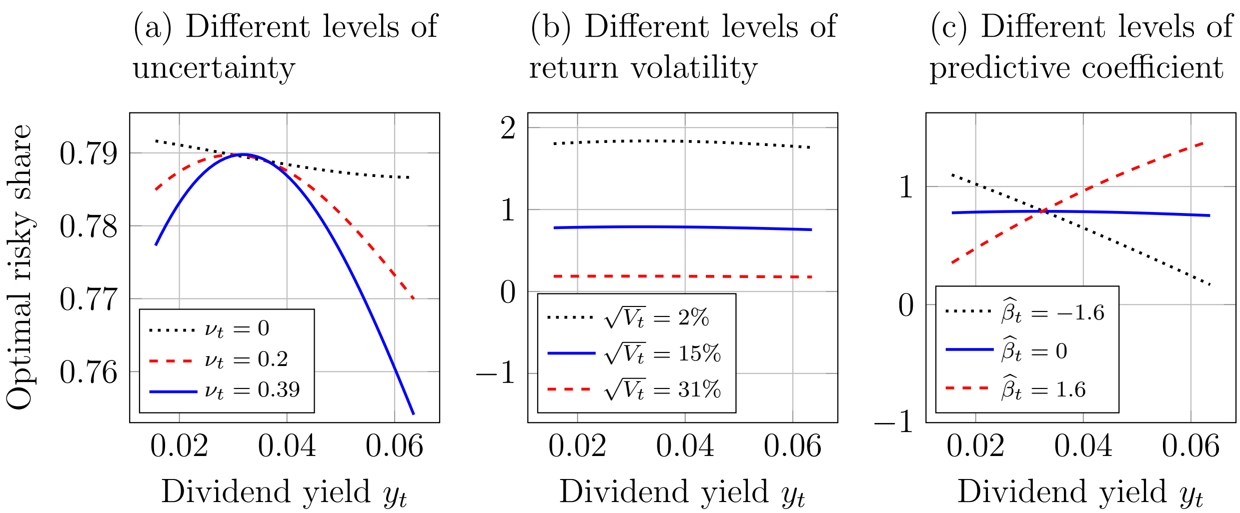

Figure 4: Impact of the dividend yield on the risky investment share.
The figure depicts the relation between the risky investment share and the dividend yield. We plot three curves corresponding to three different levels of uncertainty in panel (a), return volatility in panel (b), and the predictive coefficient in panel (c). If not stated otherwise, state variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 0.15$, and $\nu_t = \nu_{ss} = 0.39$. Parameter values are provided in Table 1.

predictive coefficient when the dividend yield is large but decreases with the predictive coefficient when the dividend yield is small (panel c).

Panel (a) of Figure 4 further shows that, in the presence of uncertainty, the risky investment share is a hump-shaped function of the dividend yield. This effect is driven by the hedging demand of the investor:

$$H_t^* \equiv w_t^* - \frac{\widehat{\mu}_t - r}{\gamma V_t}, \tag{42}$$

which reflects investor's willingness to hedge against variations in expected returns and in volatility. As Brennan (1998) and Xia (2001) demonstrate, in the presence of uncertainty, the hedging demand is negative when the investor is more risk-averse than the log-utility investor.[21]

To understand the hump-shaped pattern (which is present in all three panels, although more pronounced in panel (a) because of the scaling), note that the volatility of expected returns increases when the dividend yield moves further away from its long-term mean (see

[21]More precisely, the hedging demand consists of the sum of two negative hedging components: the expected return hedging component and the return volatility hedging component. A decrease in returns (i.e. $\frac{dP_t}{P_t} - \widehat{\mu}_t dt < 0$) implies a downward revision in expected returns, because the investor's expectations formation is "extrapolative" (Brennan, 1998), which explains the negative expected return hedging component. A decrease in returns implies an increase in return volatility because $\rho_{PV} < 0$, which explains the negative volatility hedging component.

Equation 19). This implies a stronger negative hedging component when the dividend yield is either far below or far above its long-term mean.

Finally, note that the hump shape is asymmetric in the dividend yield, i.e., the risky investment share is lower when the dividend yield reaches high values. This asymmetry is once again implied by our calibration, which specifies a negative correlation between the dividend yield and realized returns. This negative correlation implies that high dividend yields trigger a large and volatile hedging demand, but in the same time a high attention to news, as shown in Section 3.1. In other words, there is a positive relation between investor's attention and the volatility of the hedging demand.

3.3 Cost of Ignoring News

To quantify the benefits associated with paying attention to news, we compute the wealth certainty equivalent of the optimal strategy relative to that obtained when ignoring news (Xia, 2001; Das and Uppal, 2004; Liu et al., 2010). That is, the cost of ignoring news is defined as the additional fraction of wealth required by an investor who ignores the news signal (or equivalently, an investor who faces an infinite information cost) to reach the expected utility of an investor who optimally pays attention to news.

Table 2 reports the cost of ignoring news for different values of risk aversion γ and information cost k. The overall message is that the cost of ignoring news is non-negligible, reaching as much as 8.7% of wealth, especially when both risk aversion and the information cost are small. The lower the risk aversion, the larger the risky investment share, which prompts the investor to pay more attention to news (see Figure 2) and therefore increases the cost of ignoring news. The lower the cost parameter, the more valuable information becomes, wich increases the cost of ignoring news. Finally, it is worth noting that irrespective of the value of the cost parameter, a highly risk averse investor (e.g. $\gamma \approx 20$) enjoys only a marginal benefit from paying attention to news because most of her wealth is invested in the risk-free asset.

4 Empirical Analysis

Our theoretical model predicts that investors' attention is a U-shaped function of the dividend yield, a (weakly) decreasing function of stock-return variance, and a (quadratically) increasing function of uncertainty. Furthermore, the risky investment share is a hump-shaped function of the dividend yield and a decreasing function of both the stock-return variance and uncertainty. The purpose of this section is to empirically test these theoretical predictions.

Cost parameter $k \rightarrow$ **Risk aversion** $\gamma \downarrow$	10^{-2}	10^{-4}	10^{-6}
1	21.75	337.04	872.89
3	1.19	27.48	82.10
5	0.64	18.41	62.14

Table 2: Cost of ignoring news (in bps).

The cost of ignoring news represents the additional fraction of wealth required by an investor who ignores the news signal to reach the expected utility of an investor who optimally pays attention to news. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $y_t = \bar{y} = 0.04$, $\sqrt{V_t} = \sqrt{\bar{V}} = 0.15$, and $\nu_t = \nu_{ss} = 0.39$. Parameter values are provided in Table 1.

The following testable hypotheses summarize the theoretical predictions of the model:

Hypothesis 1. *At the optimum, investors' attention is a U-shaped function of the dividend yield, a (weakly) decreasing function of stock-return variance, and a (quadratically) increasing function of uncertainty.*

Hypothesis 2. *At the optimum, the risky investment share is a hump-shaped function of the dividend yield, a decreasing function of stock-return variance, and a decreasing function of uncertainty.*

To test these two hypotheses, we run regressions of the following form:

$$\text{Attention}_t = a + b_1 y_t + b_2 y_t^2 + b_3 \text{Var}_t + b_4 \text{Uncertainty}_t + b_5 \text{Uncertainty}_t^2 \quad (43)$$

$$\text{RI}_t = \bar{a} + \bar{b}_1 y_t + \bar{b}_2 y_t^2 + \bar{b}_3 \text{Var}_t + \bar{b}_4 \text{Uncertainty}_t, \quad (44)$$

where y_t is the dividend yield, Var_t is the return variance, and RI_t is our proxy for the risky investment share.

According to Hypothesis 1, we expect the regression coefficients to satisfy

$$b_2 > 0;\ |b_2| >> |b_1|;\ b_3 < 0;\ b_4 > 0 \text{ if } b_5 \text{ is set to } 0. \quad (45)$$

$$b_2 > 0;\ |b_2| >> |b_1|;\ b_3 < 0;\ b_4 = 0;\ b_5 > 0 \text{ if } b_5 \text{ is estimated}. \quad (46)$$

According to Hypothesis 2, we expect the regression coefficients to satisfy

$$\bar{b}_2 < 0;\ |\bar{b}_2| >> |\bar{b}_1|;\ \bar{b}_3 < 0;\ \bar{b}_4 < 0. \quad (47)$$

Our empirical analysis is based on the following dataset: the S&P 500 dividend yield, the variance of stock returns obtained by fitting a GARCH(1,1) model (Bollerslev, 1986) to

	Institutional attention	Institutional attention	Institutional risky investments
Intercept	0.043	0.248	−0.089**
	(0.75)	(0.11)	(−2.21)
y_t	−1.962***	−1.931***	1.308***
	(−3.55)	(−3.38)	(2.71)
y_t^2	11.582***	11.504***	−12.849***
	(2.94)	(2.96)	(−3.74)
Var_t	−25.481***	−26.875***	−1.036
	(−5.06)	(−4.19)	(−0.42)
Uncertainty_t	0.106	−0.441	0.079
	(1.24)	(−1.00)	(1.59)
Uncertainty_t^2		0.355	
		(1.089)	
Adj. R^2	0.025	0.025	0.297
N	446	446	446

Table 3: Attention and risky investments against dividend yield, return variance, and uncertainty.
The second and third columns depict the empirical relation between attention and the state variables. The fourth column depicts the empirical relation between risky investments and the state variables. Newey and West (1987) t-statistics are reported in brackets and statistical significance at the 10%, 5%, and 1% levels is labeled with */**/***, respectively. Data are at the monthly frequency from November 1977 to December 2014.

the demeaned S&P 500 returns,[22] and uncertainty as proxied by the 1-month-ahead macro-uncertainty index constructed by Jurado, Ludvigson, and Ng (2015).

Motivated by Barber and Odean (2008) and Hou, Peng, and Xiong (2009), we proxy the attention of institutional investors with the growth rate of the S&P 500 trading volume.[23] The risky investment share is proxied by the negative of the growth rate of the "Institutional Money Funds" index, which is obtained from the Federal Reserve Bank of St. Louis. Our dataset spans the period from November 1977 to December 2014 at the monthly frequency.

The second and third columns of Table 3 present the results associated to Hypothesis 1. The data confirm a U-shaped relation between institutional investors' attention and the dividend yield. The quadratic coefficient is positive, significant, and substantially larger than the linear coefficient in absolute terms. There is a negative relation between attention and stock-return variance, and a positive relation between attention and uncertainty. The signs of these coefficients are correctly predicted by the model. The negative relation between

[22]The conditional mean return is the fitted value obtained from a regression of the 1-month-ahead return on today's dividend yield.

[23]Further empirical findings by Chordia and Swaminathan (2000), Lo and Wang (2000), and Gervais, Kaniel, and Mingelgrin (2001) suggest that trading volume is a reasonable proxy for investors' attention.

attention and variance is highly significant, whereas the positive relation between attention and uncertainty is statistically insignificant. Adding the squared uncertainty to the regression has no influence on the results. Indeed, attention remains a significant U-shaped function of the dividend yield, a significant decreasing function of stock-return variance, and an insignificant function of uncertainty.

Turning now to Hypothesis 2, the fourth column of Table 3 illustrates the empirical relation between risky investments and the state variables. As predicted by the model, the relation between risky investments and the dividend yield is hump-shaped. The quadratic coefficient is negative, significant, and large compared to the linear coefficient in absolute terms. Moreover, risky investments decrease with the stock-return variance but the relation is insignificant. Finally, risky investments increase insignificantly with uncertainty, whereas the model predicts a negative relation.

Overall, all coefficients but $\bar{b}_4$ satisfy the properties described in Hypotheses 1 and 2. That is, the model seems to provide a realistic description of the observed relation between attention, risky investments, and the state variables.

5 Conclusion

This paper aims at understanding the dynamic attention behavior observed in financial markets. In most of the existing literature, investors acquire information passively in the sense that they do not control the quality of information they collect. In contrast, we consider an investor who can, at each point in time, improve the accuracy of acquired information at a cost.

Our analysis provides several interesting insights. The optimal level of attention paid to news is a U-shaped function of the stock return predictor, an increasing function of uncertainty about predictability, and a decreasing function of stock-return volatility. The former two effects reinforce each other. That is, attention is at its highest level when uncertainty is high and the stock return predictor is far away from its long-term mean. We also show that more risk averse investors spend less resources on information acquisition, and that highly persistent return predictors prompt investors to pay more attention to news.

Our analysis can be useful in a multiple asset setting, in order to understand the impact of costly dynamic information acquisition on diversification. It is also important to understand the impact of the optimal choice of attention on the equilibrium risk-free rate, equity premium, and equity return volatility. Finally, our analysis can be applied to a production economy, in order study the impact of costly dynamic information acquisition on the dynamics of aggregate consumption.

References

Abel, A., J. Eberly, and S. Panageas (2007). Optimal inattention to the stock market. *American Economic Review 97*, 244–249.

Abel, A. B., J. C. Eberly, and S. Panageas (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica 81*(4), 1455–1481.

Ait-Sahalia, Y. and M. W. Brandt (2001). Variable selection for portfolio choice. *The Journal of Finance 56*(4), 1297–1351.

Andrei, D. and M. Hasler (2015). Investor attention and stock market volatility. *Review of Financial Studies*.

Andries, M. and V. Haddad (2014). Information aversion. *Working Paper*.

Ang, A. and G. Bekaert (2007). Stock Return Predictability: Is it There? *The Review of Financial Studies 20*(3), 651–707.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Barberis, N. (2000, 02). Investing for the long run when returns are predictable. *Journal of Finance 55*(1), 225–264.

Black, F. (1976). Studies of stock price volatility changes. *American Statistical Association, Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Blackwell, D. (1953, 06). Equivalent comparisons of experiments. *Ann. Math. Statist. 24*(2), 265–272.

Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Brandt, M. W. (1999, October). Estimating portfolio and consumption choice: A conditional euler equations approach. *Journal of Finance 54*(5), 1609–1645.

Brandt, M. W., A. Goyal, P. Santa-Clara, and J. R. Stroud (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies 18*(3), 831–873.

Brennan, M. J. (1998). The role of learning in dynamic portfolio decisions. *European Finance Review 1*(3), 295–306.

Buraschi, A. and P. Whelan (2012). Term structure models with differences in beliefs. *Working Paper*.

Campbell, J. Y. and L. M. Viceira (1999, May). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics 114*(2), 433–495.

Chacko, G. and L. M. Viceira (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. *Review of Financial Studies 18*(4), 1369–1402.

Chen, H., N. Ju, and J. Miao (2014). Dynamic asset allocation with ambiguous return predictability. *Review of Economic Dynamics 17*(4), 799 – 823.

Chordia, T. and B. Swaminathan (2000). Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance 55*(2), 913–935.

Christie, A. A. (1982, December). The stochastic behavior of common stock variances : Value, leverage and interest rate effects. *Journal of Financial Economics 10*(4), 407–432.

Christoffersen, P., K. Jacobs, and K. Mimouni (2010). Volatility dynamics for the s&p500: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies 23*(8), 3141–3189.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

Cremer, J. (1982, August). A simple proof of Blackwell's "comparison of experiments" theorem. *Journal of Economic Theory 27*(2), 439–443.

Da, Z., J. Engelberg, and P. Gao (2011, October). In search of attention. *Journal of Finance 66*(5), 1461–1499.

Das, S. R. and R. Uppal (2004). Systemic risk and international portfolio choice. *The Journal of Finance 59*(6), 2809–2834.

Detemple, J. (1986). Asset pricing in a production economy with incomplete information. *The Journal of Finance 41*(2), pp. 383–391.

Detemple, J. and R. Kihlstrom (1987). Acquisition d'information dans un modèle intertemporel en temps continu. *L'Actualité économique 63*, 118–137.

Dothan, M. U. and D. Feldman (1986, June). Equilibrium interest rates and multiperiod bonds in a partially observable economy. *Journal of Finance 41*(2), 369–82.

Dumas, B., A. Kurshev, and R. Uppal (2009). Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility. *The Journal of Finance 64*(2), 579–629.

Galai, D. and O. Sade (2006). The "Ostrich Effect" and the Relationship between the Liquidity and the Yields of Financial Assets. *The Journal of Business 79*(5), 2741–2759.

Gennotte, G. (1986, July). Optimal portfolio choice under incomplete information. *Journal of Finance 41*(3), 733–46.

Gervais, S., R. Kaniel, and D. H. Mingelgrin (2001). The high-volume return premium. *The Journal of Finance 56*(3), 877–919.

Goyal, A. and I. Welch (2008, July). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies 21*(4), 1455–1508.

Grossman, S. J., R. E. Kihlstrom, and L. J. Mirman (1977, October). A Bayesian Approach to the Production of Information and Learning by Doing. *Review of Economic Studies 44*(3), 533–47.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*(2), 327–343.

Hou, K., L. Peng, and W. Xiong (2009). A tale of two anomalies: The implications of investor attention for price and earnings momentum. *Working Paper*.

Huang, L. and H. Liu (2007, 08). Rational inattention and portfolio selection. *Journal of Finance 62*(4), 1999–2040.

Judd, K. L. (1998). *Numerical Methods in Economics.* MIT Press.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2009, October). Rational attention allocation over the business cycle. NBER Working Papers 15450, National Bureau of Economic Research, Inc.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance 51*(2), pp. 385–424.

Karlsson, N., G. Loewenstein, and D. Seppi (2009, April). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty 38*(2), 95–115.

Kihlstrom, R. E. (1984). A bayesian exposition of blackwell's theorem on the comparison of experiments. *Bayesian models of economic theory*, 13–31.

Kim, T. S. and E. Omberg (1996). Dynamic nonmyopic portfolio behavior. *Review of Financial Studies 9*(1), 141–61.

Liptser, R. S. and A. N. Shiryaev (2001). *Statistics of Random Processes II.* Springer Verlag, New York.

Liu, J. (2007). Portfolio selection in stochastic environments. *Review of Financial Studies 20*(1), 1–39.

Liu, J., E. Peleg, and A. Subrahmanyam (2010, October). Information, expected utility, and portfolio choice. *Journal of Financial and Quantitative Analysis 45*(05), 1221–1251.

Lo, A. and J. Wang (2000). Trading volume: definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies 13*(2), 257–300.

Lustig, H. and A. Verdelhan (2012). Business cycle variation in the risk-return trade-off. *Journal of Monetary Economics 59*(S), S35–S49.

Marschak, J. and K. Miyasawa (1968). Economic comparability of information systems. *International Economic Review 9*(2), pp. 137–174.

Merton, R. C. (1969, August). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *The Review of Economics and Statistics 51*(3), 247–57.

Merton, R. C. (1971, December). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*(4), 373–413.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory 145*(5), 1837 – 1864.

Mondria, J. and C. Quintana-Domeque (2013). Financial contagion and attention allocation. *The Economic Journal 123*(568), 429–454.

Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

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

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *The Journal of Political Economy 111*(6), 1183–1219.

Sicherman, N., G. Loewenstein, D. J. Seppi, and S. Utkus (2015). Financial attention. *Review of Financial Studies*.

Sims, C. A. (2003, April). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

van Binsbergen, J. H. and R. S. J. Koijen (2010). Predictive regressions: A present-value approach. *The Journal of Finance 65*(4), 1439–1471.

Veldkamp, L. (2006a). Media frenzies in markets for financial information. *American Economic Review 96(3)*, 577–601.

Veldkamp, L. L. (2006b, 07). Information markets and the comovement of asset prices. *Review of Economic Studies 73*(3), 823–845.

Veronesi, P. (2000, April). How does information quality affect stock returns? *Journal of Finance 55*(2), 807–837.

Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *The Journal of Financial and Quantitative Analysis 37*(1), pp. 63–91.

Wachter, J. A. (2010). Asset allocation. *Annual Review of Financial Economics 2*(1), 175–206.

Xia, Y. (2001). Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *The Journal of Finance 56*(1), pp. 205–246.

Xiong, W. and H. Yan (2010, April). Heterogeneous expectations and bond markets. *Review of Financial Studies 23*(4), 1433–1466.

Ziegler, A. C. (2003). *Incomplete Information and Heterogeneous Beliefs in Continuous-time Finance*. Springer-Verlag Berlin Heidelberg.

A Appendix

A.1 The General Model

The correlation matrix, Λ, of the vector of Brownian motions $B \equiv (B_\beta, B_P, B_y, B_V, B_s)^\top$ is defined as

$$\Lambda \equiv \begin{pmatrix} 1 & \rho_{\beta P} & \rho_{\beta y} & \rho_{\beta V} & 0 \\ \rho_{\beta P} & 1 & \rho_{Py} & \rho_{PV} & 0 \\ \rho_{\beta y} & \rho_{Py} & 1 & \rho_{yV} & 0 \\ \rho_{\beta V} & \rho_{PV} & \rho_{yV} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{48}$$

The vector of correlated Brownian motions B can be rewritten as follows:

$$B = Chol B^{\perp}, \tag{49}$$

where $B^{\perp}$ is a 5-dimensional vector of independent Brownian motions and the matrix $Chol$ is the Cholesky decomposition of the correlation matrix Λ. The matrix $Chol$ satisfies

$$Chol = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \rho_{\beta P} & \sqrt{1-\rho_{\beta P}^2} & 0 & 0 & 0 \\ \rho_{\beta y} & \frac{\rho_{Py}-\rho_{\beta P}\rho_{\beta y}}{\sqrt{1-\rho_{\beta P}^2}} & \sqrt{1+\frac{\rho_{Py}^2-2\rho_{Py}\rho_{\beta P}\rho_{\beta y}+\rho_{\beta y}^2}{\rho_{\beta P}^2-1}} & 0 & 0 \\ \rho_{\beta V} & \frac{\rho_{PV}-\rho_{\beta P}\rho_{\beta V}}{\sqrt{1-\rho_{\beta P}^2}} & A_1 & \sqrt{1-\rho_{\beta V}^2+\frac{\left(\rho_{PV}-\rho_{\beta P}\rho_{\beta V}\right)^2}{\rho_{\beta P}^2-1}-A_1^2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{50}$$

where the coefficient A is given by

$$A_1 = \frac{\rho_{yV} - \rho_{PV}\rho_{Py} - \rho_{yV}\rho_{\beta P}^2 + \rho_{Py}\rho_{\beta P}\rho_{\beta V} + \rho_{PV}\rho_{\beta P}\rho_{\beta y} - \rho_{\beta V}\rho_{\beta y}}{\sqrt{\left(\rho_{\beta P}^2 - 1\right)\left(\rho_{Py}^2 - 1 + \rho_{\beta P}^2 - 2\rho_{Py}\rho_{\beta P}\rho_{\beta y} + \rho_{\beta y}^2\right)}}. \tag{51}$$

Following Liptser and Shiryaev (2001), the vector of state variables observed by the investor has the following dynamics

$$
\begin{aligned}
\frac{dP_t}{P_t} &= \left(\bar{\mu} + \widehat{\beta}_t(y_t - \bar{y})\right) dt + \Sigma_{P,t} d\widehat{B}_t^{\perp} && (52)\\
ds_t &= \widehat{\beta}_t dt + \Sigma_{s,t} d\widehat{B}_t^{\perp} && (53)\\
d\widehat{\beta}_t &= \lambda_\beta(\bar{\beta} - \widehat{\beta}_t)dt + \Sigma_{\beta,t} d\widehat{B}_t^{\perp} && (54)\\
dy_t &= \lambda_y(\bar{y} - y_t)dt + \Sigma_{y,t} d\widehat{B}_t^{\perp} && (55)\\
dV_t &= \lambda_V(\bar{V} - V_t)dt + \Sigma_{V,t} d\widehat{B}_t^{\perp} && (56)\\
\frac{d\nu_t}{dt} &= -A_{\nu,t}\nu_t^2 - 2B_{\nu,t}\nu_t + C_\nu, && (57)
\end{aligned}
$$

where $\widehat{\beta}$ is the filter, ν the uncertainty, and $\widehat{B}^{\perp} \equiv \left(\widehat{B}_1^{\perp}, \widehat{B}_2^{\perp}, \widehat{B}_3^{\perp}, \widehat{B}_4^{\perp}\right)^{\top}$ a 4-dimensional vector of independent and observable Brownian motions. The stochastic diffusion matrix $\Sigma \equiv (\Sigma_P, \Sigma_s, \Sigma_\beta, \Sigma_y, \Sigma_V)^{\top}$ satisfies

$$
\Sigma \equiv \begin{pmatrix} \Sigma_P \\ \Sigma_s \\ \Sigma_\beta \\ \Sigma_y \\ \Sigma_V \end{pmatrix} = \begin{pmatrix} \left(\sqrt{V} \quad 0 \quad 0 \quad 0\right) \\ \left(0 \quad 0 \quad 0 \quad \frac{1}{\sqrt{a_t}}\right) \\ \left(\frac{(y-\bar{y})\nu}{\sqrt{V}} + \sigma_\beta\rho_{\beta P} \quad \frac{1}{\sqrt{1-\rho_{Py}^2}}\left(A_2 - \frac{(y-\bar{y})\nu\rho_{Py}}{\sqrt{V}}\right) \quad \frac{1}{A_3\sqrt{V}\left(\rho_{Py}^2-1\right)}\left(A_4 + \sigma_\beta\sqrt{V}A_5\right) \quad \sqrt{a}\nu\right) \\ \sigma_y\left(\rho_{Py} \quad \sqrt{1-\rho_{Py}^2} \quad 0 \quad 0\right) \\ \sigma_V\sqrt{V}\left(\rho_{PV} \quad \sqrt{\frac{1}{1-\rho_{Py}^2}}\left(\rho_{yV} - \rho_{PV}\rho_{Py}\right) \quad A_3 \quad 0\right) \end{pmatrix}, \tag{58}
$$

where

$$
\begin{aligned}
A_2 &= \sigma_\beta\left(\rho_{\beta y} - \rho_{Py}\rho_{\beta P}\right) && (59)\\
A_3 &= \sqrt{\frac{\rho_{PV}^2 - 1 + \rho_{Py}^2 - 2\rho_{PV}\rho_{Py}\rho_{yV} + \rho_{yV}^2}{\rho_{Py}^2 - 1}} && (60)\\
A_4 &= (y - \bar{y})\,\nu\left(\rho_{PV} - \rho_{Py}\rho_{yV}\right) && (61)\\
A_5 &= \rho_{Py}^2\rho_{\beta V} - \rho_{Py}\rho_{yV}\rho_{\beta P} - \rho_{\beta V} + \rho_{yV}\rho_{\beta y} + \rho_{PV}\left(\rho_{\beta P} - \rho_{Py}\rho_{\beta y}\right). && (62)
\end{aligned}
$$

The processes A_ν and B_ν, and the parameter C_ν defining the dynamics of uncertainty satisfy

$$A_\nu = a + \frac{(y-\bar{y})^2\left(\rho_{yV}^2 - 1\right)}{VA_3^2\left(\rho_{Py}^2 - 1\right)} \tag{63}$$

$$B_\nu = \lambda_\beta - \frac{(y-\bar{y})\left(\rho_{\beta P} - \rho_{yV}^2\rho_{\beta P} - \rho_{PV}\rho_{\beta V} + \rho_{Py}\rho_{yV}\rho_{\beta V} - \rho_{Py}\rho_{\beta y} + \rho_{PV}\rho_{yV}\rho_{\beta y}\right)\sigma_\beta}{\sqrt{V}A_3^2\left(\rho_{Py}^2 - 1\right)} \tag{64}$$

$$C_\nu = -\frac{\sigma_\beta^2}{\rho_{PV}^2 - 1 + \rho_{Py}^2 - 2\rho_{PV}\rho_{Py}\rho_{yV} + \rho_{yV}^2}\left[\left(\rho_{yV}^2 - 1\right)\left(\rho_{\beta P}^2 - 1\right) - \rho_{\beta V}^2 + \rho_{Py}^2\left(\rho_{\beta V}^2 - 1\right)\right. \tag{65}$$

$$+ 2\rho_{yV}\rho_{\beta V}\rho_{\beta y} - \rho_{\beta y}^2 + 2\rho_{Py}\rho_{\beta P}\left(\rho_{\beta y} - \rho_{yV}\rho_{\beta V}\right) + 2\rho_{PV}\left(\rho_{Py}\rho_{yV} + \rho_{\beta P}\rho_{\beta V} - \left(\rho_{yV}\rho_{\beta P} + \rho_{Py}\rho_{\beta V}\right)\rho_{\beta y}\right) \tag{66}$$

$$\left. + \rho_{PV}^2\left(\rho_{\beta y}^2 - 1\right)\right]. \tag{67}$$

The investor has a CRRA utility function with risk aversion parameter γ, the subjective discount rate is δ, and the per-unit of wealth cost of paying attention is $K(a_t) = ka_t^2$, with $k \geq 0$. The investor's problem is to choose consumption c, attention to news a, and a risky investment share w so as to maximize her expected lifetime utility of consumption conditional on her information at time t, $\mathcal{F}_t$:

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) \equiv \max_{c,a,w} \mathbb{E}\left[\int_t^\infty e^{-\delta(s-t)}\frac{c_s^{1-\gamma}}{1-\gamma}ds\middle|\mathcal{F}_t\right], \tag{68}$$

subject to the budget constraint[24]

$$dW_t = \left[(r_t - ka_t^2)W_t + w_tW_t\left(\widehat{\mu}_t - r_t\right) - c_t\right]dt + w_tW_t\left[\sqrt{V_t} \quad 0 \quad 0 \quad 0\right]d\widehat{B}_t^\perp. \tag{69}$$

The optimality condition is

$$0 = -\delta J + \max_{c,a,w}\left(\frac{c^{1-\gamma}}{1-\gamma} + \mathcal{D}^X J\right), \tag{70}$$

where $\mathcal{D}^X$ is the infinitesimal generator associated to the vector of state variables $X \equiv (W, \widehat{\mu}, y, V, \nu)^\top$. Conjecturing that the value function satisfies

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t), \tag{71}$$

computing the first-order conditions of (70), and substituting them back in (70) yields a partial differential equation (PDE) for the function $\phi(\widehat{\mu}, y, V, \nu)$.

[24]Note that, because their is a one-to-one mapping between $\widehat{\beta}$ and $\widehat{\mu}$ (i.e. $\widehat{\mu} = \bar{\mu} + \widehat{\beta}(y - \bar{y})$), the state variables $\widehat{\beta}$ and $\widehat{\mu}$ are interchangeable.

A.2 Numerical Solution Method

As mentioned in Appendix A.1, substituting the first-order conditions back in Equation (70) yields a PDE the function $\phi(\widehat{\mu}, y, V, \nu)$ has to satisfy. We numerically solve this PDE using the Chebyshev collocation method described in Judd (1998). That is, we approximate the function $\phi(\widehat{\mu}, y, V, \nu)$ as follows:

$$\phi(\widehat{\mu}, y, V, \nu) \approx P\left(\widehat{\mu}, y, V, \nu\right) = \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\sum_{l=0}^{L} a_{i,j,k,l} T_i\left(\widehat{\mu}\right) T_j\left(y\right) T_k\left(V\right) T_l\left(\nu\right),$$

where T_m is the Chebyshev polynomial of order m. We mesh the roots of the Chebyshev polynomials of order $I+1$, $J+1$, $K+1$, and $L+1$ to obtain the interpolation nodes. Substituting $P\left(\widehat{\mu}, y, V, \nu\right)$ and its partial derivatives in the PDE and evaluating the latter at the interpolation nodes yields a system of $(I+1)\times(J+1)\times(K+1)\times(L+1)$ equations with $(I+1)\times(J+1)\times(K+1)\times(L+1)$ unknowns (the coefficients $a_{i,j,k,l}$) that we solve numerically.

A.3 Dynamics of the Vector of State Variables

The dynamics of the vector of state variables $Z \equiv \left[\widehat{\mu}, y, V, \nu\right]^\top$ are defined as follows:

$$dZ_t = m_t dt + \Sigma_{1,t} d\widehat{B}^{\perp}_{1,t} + \Sigma_{2,t}\begin{pmatrix} d\widehat{B}^{\perp}_{2,t} \\ d\widehat{B}^{\perp}_{3,t} \\ d\widehat{B}^{\perp}_{4,t}\end{pmatrix}, \tag{72}$$

where the 4-dimensional vector of drift m, the 4-dimensional vector of diffusion Σ_1, and the 4×3 matrix of diffusion Σ_2 satisfy

$$m = \begin{pmatrix} \left(\lambda_y + \lambda_\beta\right)\left(\bar{\mu} + \frac{\bar{\beta}\lambda_\beta(y-\bar{y})}{\lambda_y+\lambda_\beta} - \widehat{\mu}\right) \\ \lambda_y(\bar{y} - y) \\ \lambda_V(\bar{V} - V) \\ -\left(\frac{(y-\bar{y})^2}{V} + a\right)\nu^2 - 2\lambda_\beta \nu + \sigma_\beta^2 \end{pmatrix} \qquad \Sigma_1 = \begin{pmatrix} \frac{(y-\bar{y})^2\nu}{\sqrt{V}} \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{73}$$

$$\Sigma_2 = \begin{pmatrix} \sigma_y \frac{\widehat{\mu}-\bar{\mu}}{y-\bar{y}} & 0 & \nu\sqrt{a}(y-\bar{y}) \\ \sigma_y & 0 & 0 \\ 0 & \sigma_V\sqrt{V} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{74}$$