Dynamic Attention Behavior Under Return Predictability*

Daniel Andrei†  Michael Hasler‡

April 12, 2016

Abstract

We investigate the dynamic problem of how much attention an investor should pay to news in order to learn about stock-return predictability and maximize expected lifetime utility. We show that the optimal amount of attention is U-shaped in the return predictor, increasing with both uncertainty and the magnitude of the predictive coefficient, and decreasing with stock-return volatility. The optimal risky asset position exhibits a negative hedging demand, is hump-shaped in the return predictor, decreasing with uncertainty, and increasing with stock-return volatility. We test and find empirical support for these theoretical predictions.

*We would like to thank Patrick Augustin, Craig Boreth, Michael Brennan, Peter Christoffersen, Alexander David, Alexandre Jeanneret, Chayawat Ornthanalai, Alberto Rossi, and Alexey Rubtsov for insightful suggestions and comments. We are also grateful for comments received from conference/seminar participants at HEC-McGill Winter Finance Conference 2016, ISF 2015, QWAFSAFEW 2016, Ryerson University, UCLA Anderson, 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

Growing empirical evidence suggests that investor attention is time-varying. This affects trading and asset prices, and can significantly impact financial markets. The aim of this paper is to understand whether these fluctuations result from rational information gathering behavior by investors.

One segment of the literature postulates that information has a hedonic impact on utility. This can generate different levels of attention depending on the state of the world and can thus explain the fluctuations we observe. There is evidence in favor of this line of research. Another part of the literature postulates that investors have limited ability to process information (Sims, 2003). Therefore, information does not directly enter the utility function, but indirectly helps investors make better decisions under uncertainty.

We follow an alternative approach, excluding hedonic effects or information processing constraints. Instead, we assume that information is costly. Our setup is based on the incomplete-information literature which started with seminal papers by Gennotte (1986), Detemple (1986), and Dothan and Feldman (1986). One implicit assumption made by most of this literature is that the set of available information is exogenously determined. The objective of this paper is to relax this assumption, to bring our theoretical predictions to data, and to understand whether this theory can explain the observed fluctuations in investors’ attention.

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

---

1Barber and Odean (2008); Da, Engelberg, and Gao (2011); Sicherman, Loewenstein, Seppi, and Utkus (2015); Fisher, Martineau, and Sheng (2016).
2Chien, Cole, and Lustig (2012); Andrei and Hasler (2015); Fisher et al. (2016).
4Sicherman et al. (2015).
5See also Veldkamp (2006b,a); Van Nieuwerburgh and Veldkamp (2006, 2010); Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016).
The investor assumes that expected returns are a linear function of an observable predictive variable (Xia, 2001). The focus of the paper is on the predictive coefficient, which is stochastic and unobservable. The investor uses all of the available historical data to estimate the predictive coefficient and then construct a forecast of returns. In addition to historical data, the investor can also choose to improve her estimation by collecting news, but at a cost. Because perfect learning—observing expected returns—has an infinite cost, the investor must choose an optimal, finite amount of attention. We characterize this optimal amount of attention and its responsiveness to changes in the state variables of the model.

The predictive variable determines, to a large extent, the optimal amount of attention to news. We show that the investor pays more attention to news the further away the predictive variable is from its long-term mean. In these states, the investor attempts to profit from the reversion to the mean of expected returns and thus acquires additional information to bet on the upcoming trend. Because this arises whenever the predictive variable is far away from its long-term mean, attention exhibits a U-shaped pattern.

Furthermore, we show that greater uncertainty unambiguously increases the attention to news. The reason is that greater uncertainty increases the volatility of expected returns and therefore increases the likelihood of larger future trends. Since the investor can efficiently exploit these trends only if her estimate of the predictive coefficient is accurate, the optimal decision is to pay more attention to news.

We also find that attention to news decreases as a function of the volatility of realized returns, which we assume to be stochastic (Chacko and Viceira, 2005; Liu, 2007). When the volatility of realized returns increases, the quality of information provided by those returns deteriorates, which decreases the volatility of expected returns. This, in turn, decreases the likelihood of large future trends and prompts the investor to decrease her attention to news. To our knowledge, this prediction of decreased attention during periods of high realized-return volatility has not been postulated before.

Finally, we show that attention to news unambiguously increases when the estimated predictive coefficient is large in magnitude. This situation is characterized by highly volatile expected returns and prompts the investor to be more attentive to news.

The optimal risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand, because expected returns are positively correlated to

---

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

8 Note that in our model there is a distinction between “expected return volatility” and “realized return volatility.” The former depends on all state variables and can be thought of as a forward looking measure (which can be proxied by the VIX index), whereas the latter is a contemporaneous measure. As such, our model predicts that an investors’ attention to news increases when the expected return volatility is high (see also Fisher et al., 2016) but decreases when the realized return volatility is high.
returns. The hedging demand is hump-shaped in the predictive variable, and its magnitude increases with uncertainty and decreases with the stock-return volatility. This is because expected returns are particularly sensitive to return shocks when the predictive variable is far from its long-term mean, uncertainty is high, and stock-return volatility is low.

In the empirical section of the paper, we test the dependence of attention to news and the risky investment share on the state variables of our model. We first calibrate the model to S&P 500 returns and define the predictive variable as the S&P 500 earnings-to-price ratio. We then build three model-implied time series: uncertainty, attention, and risky investment share. We compare these time series with their empirical counterparts and find a strong positive correlation in each case. We find that the model-implied attention to news is counter-cyclical, and strongly predicts the VIX index. Consistent with the predictions of the model, we find that the empirical proxy for attention to news is indeed U-shaped in the predictive variable, it increases with both uncertainty and the magnitude of the predictive coefficient, and it decreases with stock-return volatility. We discuss additional empirical evidence by Fisher et al. (2016) that supports the U-shaped prediction. In addition, we show that the empirical proxy for the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand whose shape mirrors the model’s prediction. We therefore conclude that our setup accurately describes investors’ dynamic asset and attention allocation behavior.

This paper complements a large body of literature which considers portfolio selection problems with stochastic expected returns, stochastic volatility, incomplete information, and uncertainty about predictability. In the literature on costly information acquisition, Detemple and Kihlstrom (1987) analyze the demand for, and equilibrium price of information in the context of a production economy, and Huang and Liu (2007) consider a static information acquisition problem. In contrast, we consider a dynamic information acquisition problem in the presence of uncertainty about stock-return predictability. Veldkamp (2006a,c) shows that costly information acquisition helps explain excess co-movement and the simultaneous increases in emerging markets’ media coverage and equity prices, whereas Kacperczyk et al. (2016) study the problem of investment managers who choose how to allocate a fixed amount of attention to aggregate versus stock-specific information. In our model, the total flow of information is not bounded by a capacity constraint outside of the model, but is endogenously

---

determined through an optimization exercise performed by the investor. Our study sheds light on the dynamic relationship between attention, investments, and relevant state variables, namely expected returns, uncertainty, and volatility. We provide both a theoretical and an empirical analysis, which help explicate these fundamental relations.

The remainder of the paper is organized as follows. Section 2 describes the economy and examines the optimal attention, consumption, and portfolio choice problem of the investor. Section 3 calibrates the parameters of the model and describes the results. Section 4 performs the empirical analysis. Section 5 concludes. The Appendix contains all proofs.

2 The model

Consider an economy populated by an investor with utility function defined by

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

where $c_t$ is the consumption at time $t$, $\delta$ 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 risk-free asset paying a constant rate of return $r^f$, and one risky asset (the stock) whose price dynamics satisfy

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

where $\mu_t$ is the instantaneous expected return on the stock and $V_t$ is the instantaneous variance of stock returns.

The investor operates under partial knowledge of the economy. Specifically, the expected return $\mu_t$ is unobservable (Brennan, 1998; Xia, 2001; Ziegler, 2003), but the investor knows that it is an affine function of an observable state variable $y_t$. That is, the expected return satisfies

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

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

\[ dy_t = \lambda_y (\bar{y} - y_t) dt + \sigma_y dB_{y,t} \]  
\[ d\beta_t = \lambda_\beta (\bar{\beta} - \beta_t) dt + \sigma_\beta dB_{\beta,t}, \]  

where we assume that \( \bar{y}, \lambda_y, \sigma_y, \bar{\beta}, \lambda_\beta, \text{ and } \sigma_\beta \) are known constants. The variance of stock returns is observable and follows a square-root process (Heston, 1993; Liu, 2007):

\[ dV_t = \lambda_V (\bar{V} - V_t) dt + \sigma_V \sqrt{V_t} dB_{V,t}, \]  

where \( \bar{V}, \lambda_V \) and \( \sigma_V \) are known constants. The four Brownian motions \( B_{P,t}, B_{y,t}, B_{\beta,t}, \text{ and } B_{V,t} \) are independent from each other.

### 2.1 The inference process: active learning

Given the dynamics of the state variables described above, the investor’s problem consists of inferring the predictive coefficient \( \beta_t \) before choosing an optimal portfolio and consumption rule that maximizes the expected lifetime utility of consumption.

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

\[ ds_t = \beta_t dt + \frac{1}{\sqrt{a_t}} dB_{s,t}, \]  

where \( B_{s,t} \) is an Brownian motion independent of \( B_{P,t}, B_{y,t}, B_{\beta,t}, \text{ and } B_{V,t} \).

The dynamics of the news signal (7) are interpreted as follows. Assume the investor acquires \( n_t \) signals of equal precision \( s^j_t, j = 1, \ldots, n_t \) at time \( t \):

\[ ds^j_t = \beta_t dt + \sigma_s dB^j_{s,t}, \quad j = 1, \ldots, n_t \]  

where \( B^j_{s,t} \) is independent of \( B_{P,t}, B_{y,t}, B_{\beta,t}, \text{ and } B_{V,t} \) for all \( j \) and \( B^j_{s,t} \perp B^i_{s,t}, \forall j \neq i. \)

Aggregating yields the following dynamics of the aggregate signal \( s_t \) acquired by the investor

\[ ds_t \equiv \frac{1}{n_t} \sum_{j=1}^{n_t} ds^j_t = \beta_t dt + \frac{\sigma_s}{\sqrt{n_t}} dB_{s,t}, \]  

where \( B_{s,t} \) is independent from \( B_{P,t}, B_{y,t}, B_{\beta,t}, B_{V,t} \). Setting \( \frac{\sigma_s}{\sqrt{n_t}} \equiv \frac{1}{\sqrt{a_t}} \) in Equation (9) leads to Equation (7). That is, the investor controls the accuracy \( a_t \) of the aggregate signal by...
choosing the number of individual signals \( n_t \) she acquires. When the investor is attentive to news, the number of individual signals she acquires is large and the aggregate signal is accurate. When the investor is inattentive to news, the number of individual signals she acquires is small and the aggregate signal is inaccurate. Given this, we call \( a_t \) the investor’s attention to news.

Denote by \( \mathcal{F}_t \) the information set of the investor at time \( t \). This information set 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.

Let us denote by \( \hat{\beta}_t \equiv \mathbb{E}[\beta_t|\mathcal{F}_t] \) the estimated predictive coefficient and its posterior variance by \( \nu_t \equiv \mathbb{E}[(\beta_t - \hat{\beta}_t)^2|\mathcal{F}_t] \). Thus,

\[
\beta_t \sim N(\hat{\beta}_t, \nu_t),
\]

where \( N(m, v) \) denotes the Normal distribution with mean \( m \) and variance \( v \). Henceforth, we refer to the estimated predictive coefficient \( \hat{\beta}_t \) and the posterior variance \( \nu_t \) 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

\[
\frac{dP_t}{P_t} = \left( \bar{\mu} + \hat{\beta}_t(y_t - \bar{y}) \right) dt + \left[ \sqrt{V_t} \ 0 \ 0 \ 0 \right] d\hat{B}_t^t
\]

\[
ds_t = \hat{\beta}_t dt + \left[ 0 \ 0 \ 0 \ \frac{1}{\sqrt{\nu_t}} \right] d\hat{B}_t^t
\]

\[
dy_t = \lambda_y(y_t - \bar{y}) dt + \left[ 0 \ \sigma_y \ 0 \ 0 \right] d\hat{B}_t^t
\]

\[
dV_t = \lambda_V(V_t - \bar{V}) dt + \left[ 0 \ 0 \ \sigma_V \sqrt{V_t} \ 0 \right] d\hat{B}_t^t.
\]

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

\[
d\hat{\beta}_t = \lambda_{\beta}(\bar{\beta} - \hat{\beta}_t) dt + \left[ \frac{\nu_t(y_t - \bar{y})}{\sqrt{\nu_t}} \ 0 \ \nu_t \sqrt{a_t} \right] d\hat{B}_t^t
\]

\[
d\nu_t dt = - \left( \frac{(y_t - \bar{y})^2}{V_t} + a_t \right) \nu_t^2 - 2\lambda_{\beta}\nu_t + \sigma_{\beta}^2,
\]
where $\hat{B}_t^\perp \equiv \left[ \hat{B}_{1,t}^\perp, \hat{B}_{2,t}^\perp, \hat{B}_{3,t}^\perp, \hat{B}_{4,t}^\perp \right]^\top$ is a 4-dimensional vector consisting of independent Brownian motions observable under investor’s filtration.

**Proof.** See Liptser and Shiryaev (2001).

Equations (15) and (16) describe the 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 (15) shows, the investor assigns stochastic weights to these four sources of information. As we will describe below, the size of these weights depend on the relative informativeness of each source of information.

Equation (16) describes the change in uncertainty when the investor controls her attention to news. Uncertainty is locally deterministic and decreases faster when the investor’s attention is high. The decline in uncertainty is weaker when the predictive coefficient is more persistent (i.e. low $\lambda_\beta$) or when the volatility of realized returns $V_t$ is high. Finally, the last term in Equation (16) shows that the larger the volatility of the predictive coefficient, the stronger the increase in uncertainty over time.

The informativeness of the signal depends on the investor’s attention, which impacts learning in two ways. First, it has a direct impact on the instantaneous volatility of the filter in Equation (15). Second, it drives the drift of uncertainty in Equation (16). We analyze each of these two effects separately. To facilitate our discussion, we refer to $d\hat{B}_{1,t}^\perp$ as return shocks and to $d\hat{B}_{4,t}^\perp$ 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 $\nu_t$, 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 insight on how the investor updates her beliefs using each piece of information.

Suppose that $y_t > \bar{y}$. Then, an unexpectedly high return ($d\hat{B}_{1,t}^\perp > 0$) means that the current estimate of $\beta_t$ is too low, and the investor adjusts $\hat{\beta}_t$ upwards. The opposite happens when $y_t < \bar{y}$: An unexpectedly high return means that the current estimate of $\beta_t$ is too high, and the investor adjusts $\hat{\beta}_t$ 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).

10Because 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 $\beta_t$. Hence, the second and third components of the diffusion of $\hat{\beta}_t$ are both equal to zero.
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.

Overall, the instantaneous variance of the filter is an increasing function of attention:

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

(17)

As attention converges to infinity, the instantaneous variance of the filter converges to its upper bound \(\sigma_{\beta}^2\). This upper bound represents the variance of the filter when the predictive coefficient \(\beta_t\) is perfectly observable.

### 2.1.2 The impact of attention on uncertainty

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 \(\beta_t\) 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 \(\beta_t\) and 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:

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

(18)

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

$$d\hat{\mu}_t = (\lambda_y + \lambda_{\beta}) \left( \mu + \frac{\beta \lambda_{\beta}(y_t - \bar{y})}{\lambda_y + \lambda_{\beta}} - \hat{\mu}_t \right) dt + \left[ \frac{(y_t - \bar{y})^2}{\sigma_y \sqrt{V_t}} \right] d\hat{B}_t^\perp.$$

(19)

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

\[
\eta_t \equiv \mathbb{E} \left[ (\mu_t - \hat{\mu}_t)^2 \mid \mathcal{F}_t \right] = (y_t - \bar{y})^2 \nu_t. \tag{20}
\]

The instantaneous variance of the estimated expected return is

\[
\text{Var}[d\hat{\mu}_t] = \nu_t^2 (y_t - \bar{y})^2 \left( a_t + \frac{(y_t - \bar{y})^2}{V_t} \right) + \sigma_y^2 \hat{\beta}_t^2. \tag{21}
\]

\( \text{Var}[d\hat{\mu}_t] \) is a monotone increasing function of attention. Its maximum depends on both \( \hat{\beta}_t \) and \( y_t \) and is given by

\[
\lim_{a_t \to \infty} \text{Var}[d\hat{\mu}_t] = \sigma_y^2 (y_t - \bar{y})^2 + \sigma_y^2 \hat{\beta}_t^2. \tag{22}
\]

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 \( \hat{\beta} \) 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 (19), when the filter \( \hat{\beta}_t \) is large, expected returns react strongly to changes in the predictive variable \( y_t \) (the second component of the diffusion). Furthermore, if investor’s attention is high, expected returns react strongly to news shocks, but only when \( y_t \neq \bar{y} \) (the fourth component of the diffusion). This concurs with the relation between returns and the predictive variable described in Equation (3), whereby more news on the predictive coefficient \( \beta_t \)—no matter how accurate—is not going to change investor’s view about expected returns if \( y_t - \bar{y} = 0 \).

High uncertainty \( \nu_t \) magnifies the sensitivity of expected returns to return shocks (the first component of the diffusion). When \( y_t - \bar{y} \neq 0 \), an increase in uncertainty increases the variance of expected returns. This is shown in Equation (21). 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 (22).\(^{11}\)

To summarize, attention to news 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. At the same time, more attention yields lower future uncertainty by magnifying the negative component of its drift.

\(^{11}\)We derive the latter equation by applying Itô’s lemma to the expected return in Equation (3) and by assuming that the predictive coefficient \( \beta_t \) is perfectly observable.
2.3 Optimal attention, portfolio choice, and consumption

Turning now to the investor’s optimization problem, we consider $\hat{\mu}_t$ as a state variable instead of $\hat{\beta}_t$, with the aim to better interpret and characterize our results. 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

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

subject to the budget constraint

$$dW_t = [(r^f - K(a_t))W_t + w_tW_t (\hat{\mu}_t - r^f) - c_t] dt + w_tW_t \begin{bmatrix} \sqrt{V_t} & 0 & 0 \end{bmatrix} d\hat{B}_t^t.$$  (24)

To reflect the fact that prices and costs tend to increase as time passes, we follow Bansal and Shaliastovich (2011) and assume that the total information cost, $W_t K(a_t)$, is linear in wealth. This assumption preserves the homogeneity of the value function in wealth and therefore implies that attention is stationary at the optimum.\(^\text{12}\) 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 be inattentive to news ($a_t = 0$), and therefore to learn using only the information provided by the price $P_t$, the predictive variable $y_t$, and the variance of stock returns $V_t$, 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.

**Proposition 2.** The optimal consumption $c^*_t$, risky investment share $w^*_t$, and attention to

\(^{12}\)Liu, Peleg, and Subrahmanyam (2010) show that the more wealthy the investor is, the more valuable (in dollar terms) information becomes. The reason is that signals provide information on returns and not on growth in dollar terms. Our specification of the information cost implies that a piece of information of a given quality rewards the investor equivalently (in net of information costs return terms) today and at any future point in time.
news $a_t^*$ are given by

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

(25)

$$w_t^* = \frac{\hat{\mu}_t - rf}{V_t} - \frac{J_W}{J_{WW}W_t} \left( \frac{\nu_t(y_t - \bar{y})^2}{V_t} \frac{-J_{W\mu}}{J_{WW}W_t} \right)$$

(26)

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

(27)

where $\Phi(\cdot)$ is a positive and increasing function defined as the inverse of the derivative of the cost function, $\Phi(\cdot) \equiv K'(\cdot)^{-1}$.

**Proof.** See Appendix A.1.

Equation (25) is the standard optimal consumption derived by Merton (1971). The optimal risky investment share, expressed in Equation (26), comprises a myopic and a hedging portfolio (Merton, 1971). The hedging term represents the effect of parameter learning and significantly impacts 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.

Our object of focus is the optimal attention $a_t^*$, expressed in Equation (27). Since the function $\Phi(\cdot)$ is positive and increasing, we can directly analyze the term in brackets. Two factors drive the optimal level of attention: the *state risk aversion factor* $J_{\mu\mu}$, which measures the extent to which the investor (dis)likes variations in expected returns and the *uncertainty factor* $J_{\nu}$, which measures the extent to which the investor (dis)likes uncertainty. Recall from Section 2.2 that attention drives both the variance of expected returns and the drift of uncertainty. The state risk aversion factor $J_{\mu\mu}$ 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 21). The uncertainty factor $J_{\nu}$ is multiplied by $-\nu_t^2$, which is the marginal effect of attention on the drift of $\nu_t$ (see Equation 16).

Because our setup features mean-reverting expected returns, the value function is convex in $\hat{\mu}_t$, which yields $J_{\mu\mu} > 0$ (Kim and Omberg, 1996). That is, the investor prefers expected return volatility because it creates the possibility of trends that she can exploit. The higher the volatility, the higher the convexity of the value function and thus the investor pays greater attention to accurately estimate the predictive coefficient and efficiently exploit future trends.

Two opposing forces determine the sign of the uncertainty factor $J_{\nu}$. First, the investor dislikes uncertainty; second, the investor likes uncertainty because it leads to higher expected return volatility and a higher likelihood of trends. Depending on which of these two forces dominates, the uncertainty factor $J_{\nu}$ is positive or negative. If it is negative, the investor
acquires more information to reduce uncertainty. If it is positive, the investor acquires less information in order to keep expected return volatility high and take advantage of trends.

Overall, Proposition 2 implies 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. These results are independent on the investor’s utility function. As Equation (27) shows, the effects of uncertainty and the predictive variable 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.

2.4 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 $\gamma$. In addition, the cost function for attention is specified in quadratic form:

$$K(a_t) = ka_t^2, \quad (28)$$

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

$$\Phi(x) = \frac{x}{2k}. \quad (29)$$

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

$$J(W_t, \hat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1 - \gamma} \phi(\hat{\mu}_t, y_t, V_t, \nu_t). \quad (30)$$

Computing the partial derivatives of $J$ as a function of the partial derivatives of $\phi$ 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.** With CRRA utility and quadratic attention cost, the optimal consumption $c_t^*$, risky investment share $w_t^*$, and attention to news $a_t^*$ are given by

$$c_t^* = \phi^{-1/\gamma} W_t \quad (31)$$

$$w_t^* = \frac{\hat{\mu}_t - r_f}{\gamma V_t} + \frac{\nu_t (y_t - \bar{y})^2}{\gamma V_t} \frac{\phi}{\phi} \quad (32)$$

$$a_t^* = \nu_t^2 \left( \frac{\phi}{\phi} \frac{1}{2k(\gamma - 1)} + \frac{-\phi_{\mu\mu} (y_t - \bar{y})^2}{4k(\gamma - 1)} \right). \quad (33)$$
The optimal consumption defined in Equation (31) is well-known (Merton, 1971) and does not require any further analysis. The optimal risky investment share defined in Equation (32) is analyzed by Xia (2001) in a setup with constant stock-return volatility.\footnote{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 $\beta$. These correlations are set to zero in the present case. Note also that, although a similar decomposition appears in Xia (2001), the attention level affects the shape of the value function, causing differences between the portfolio holdings obtained in the two models.} We discuss the dependence of the risky investment share and its hedging components on the state variables in Section 3.3.

The optimal attention defined in Equation (33) becomes a strictly increasing quadratic function of uncertainty. We discuss the dependence of the investor’s attention on the state variables in Section 3.2.

Although an indirect dependence of attention on the variance of stock returns, $V_t$, arises through the function $\phi$, the variance of stock return has no direct impact on the investor’s attention. In Section 3.2 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 determinants of optimal attention and risky investment share. We first calibrate the parameters of the model, then we show that attention is a U-shaped function of the predictive variable, a decreasing function of the stock-return variance, and an increasing function of both the absolute value of the predictive coefficient and uncertainty. The risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand that is hump-shaped in the predictive variable. Furthermore, the size of the hedging demand increases with uncertainty and decreases with stock-return volatility.

#### 3.1 Calibration

We calibrate the parameters of the model using three datasets. First, we consider the S&P 500 earnings-to-price ratio to be the predictive variable $y_t$ of S&P 500 returns $r_t$.\footnote{Several other predictive variables have been identified. They include past market returns, the dividend yield, nominal interest rates, and expected inflation among others. See Goyal and Welch (2008) for a comprehensive survey.} This dataset is obtained from Robert Shiller’s website, at monthly frequency from 01/1950 to 12/2014. The dynamics of the predictive variable provided in Proposition 1 imply that the
long-term mean of $y_t$ is $\bar{y}$, the long-term variance of $y_t$ is $\sigma_y^2/(2\lambda_y)$, and $\text{cov}(y_{t+\Delta}, y_t)/\text{var}(y_t) = e^{-\lambda_y\Delta}$, where $\Delta = 1/12 = 1$ month. Solving these three moment conditions yields the parameters $\bar{y}$, $\sigma_y$, and $\lambda_y$.

Second, we jointly estimate the long-term expected return $\bar{\mu}$ and the time-varying predictive coefficients $\hat{\beta}_t$ by performing 60-month rolling window regressions of 1-month-ahead returns on current demeaned earnings-to-price ratios. That is, $\bar{\mu}$ and $\hat{\beta}_t$ satisfy

$$r_{t+\Delta} = \left[\bar{\mu} + \hat{\beta}_j\Delta + 60\Delta (y_{j\Delta} - \bar{y})\right] \Delta + \epsilon_{t+\Delta}, \quad t \in (j\Delta, j\Delta + 59\Delta),$$

where $j = 0, \ldots, N - 1$ is the window’s index, $N$ is the total number of windows, and $\epsilon_t$ is a random variable with mean 0 and variance $V_i \Delta$. The dynamics of the predictive coefficient provided in Proposition 1 imply that the long-term mean of $\hat{\beta}_t$ is $\bar{\beta}$, the long-term variance of $\hat{\beta}_t$ can be approximated by $\sigma_{\beta}^2/(2\lambda_{\beta})$, and $\text{cov}(\hat{\beta}_{t+\Delta}, \hat{\beta}_t)/\text{var}(\hat{\beta}_t) = e^{-\lambda_{\beta}\Delta}$. Solving these three moment conditions yields the parameters $\bar{\beta}$, $\sigma_{\beta}$, and $\lambda_{\beta}$.

Third, we compute the demeaned returns $\epsilon_t$ as follows

$$\epsilon_{j\Delta+61\Delta} = r_{j\Delta+61\Delta} - \left[\bar{\mu} + \hat{\beta}_j\Delta + 60\Delta (y_{j\Delta+60\Delta} - \bar{y})\right] \Delta,$$

and estimate their conditional variance $V_i \Delta$ by fitting a GARCH(1,1) model (Bollerslev, 1986). The dynamics of the stock return variance provided in Proposition 1 imply that the long-term mean of $V_t$ is $\bar{V}$, the long-term variance of $V_t$ is $\sigma_V^2 \bar{V}/2\lambda_V$, and $\text{cov}(V_{t+\Delta}, V_t)/\text{var}(V_t) = e^{-\lambda_V\Delta}$. Solving these three moment conditions yields the parameters $\bar{V}$, $\sigma_V$, and $\lambda_V$.

The estimated parameter values are provided in Table 1. In addition, we set the risk-free rate to its historical mean $r_f = 5.08\%$. We fix the cost parameter to $k = 0.1$, the risk aversion coefficient to $\gamma = 3$, and the subjective discount rate to $\delta = 0.01$. Using these parameters, we numerically solve the partial differential equation resulting from the specification in Equation (30) by applying the Chebyshev collocation method described in Judd (1998). More details on the solution method are provided in Appendix A.2.

### 3.2 Optimal attention

Four state variables impact the optimal attention: the predictive variable $y_t$, the uncertainty $\nu_t$, the stock-return variance $V_t$, and the predictive coefficient $\hat{\beta}_t$. Since the predictive variable is the main determinant of expected returns, we choose to plot the optimal attention against the predictive variable for different values of the other state variables.

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 benchmark solid blue
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean-reversion speed of stock-return variance</td>
<td>$\lambda_V$</td>
<td>1.9592</td>
</tr>
<tr>
<td>Long-term mean of stock-return variance</td>
<td>$\bar{V}$</td>
<td>0.0158</td>
</tr>
<tr>
<td>Volatility of stock-return variance</td>
<td>$\sigma_V$</td>
<td>0.1760</td>
</tr>
<tr>
<td>Mean-reversion speed of earning-to-price ratio</td>
<td>$\lambda_y$</td>
<td>0.1163</td>
</tr>
<tr>
<td>Long-term mean of earning-to-price ratio</td>
<td>$\bar{y}$</td>
<td>0.0686</td>
</tr>
<tr>
<td>Volatility of earning-to-price ratio</td>
<td>$\sigma_y$</td>
<td>0.0136</td>
</tr>
<tr>
<td>Long-term expected return</td>
<td>$\bar{\mu}$</td>
<td>0.0685</td>
</tr>
<tr>
<td>Mean-reversion speed of $\beta$</td>
<td>$\lambda_\beta$</td>
<td>0.2151</td>
</tr>
<tr>
<td>Long-term mean of $\beta$</td>
<td>$\bar{\beta}$</td>
<td>0.0021</td>
</tr>
<tr>
<td>Volatility of $\beta$</td>
<td>$\sigma_\beta$</td>
<td>0.9531</td>
</tr>
</tbody>
</table>

Table 1: Calibrated Parameter Values.

The three panels depict the relation between attention and the predictive variable. 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, the state variables are $\hat{\beta}_t = \beta = 0$, $\sqrt{\bar{V}_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. Parameter values are provided in Table 1.

Figure 1: Impact of the predictive variable on attention.

The three panels depict the relation between attention and the predictive variable. 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, the state variables are $\hat{\beta}_t = \beta = 0$, $\sqrt{\bar{V}_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. Parameter values are provided in Table 1.

line is obtained by setting $\nu_t$, $V_t$, and $\hat{\beta}_t$ to their long-term levels $\nu_{ss}$, $\sqrt{\bar{V}}$, and $\bar{\beta}$.\(^{15}\)

All three panels of Figure 1 confirm the intuition provided in Equation (33) that attention is a U-shaped function of the demeaned predictive variable $y_t - \bar{y}$. This is because when the predictive variable is close to its long-term mean, the investor knows that the expected return is equal to $\bar{\mu}$ and therefore has weak incentives to pay for information. In contrast, when

\(^{15}\)The “steady-state” uncertainty $\nu_{ss}$ 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 $\nu_t$, 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 predictive variable, the return variance, and the signal) are uninformative when $y_t = \bar{y}$ and $a_t = 0$.
the predictive variable is far from its long-term mean, there is a trend in expected returns that the investor can efficiently exploit, but only if the predictive coefficient is accurately estimated. The investor’s optimal reaction to this situation is to pay attention, efficiently exploit the trend, and profit from it.

Panel (a) of Figure 1 shows that uncertainty drives investor attention in two ways. First, it increases the curvature of the U-shaped relation between attention and the predictive variable through the presence of \( \nu^2_t \) in Equation (33). Second, it slightly increases the level of the U-shaped relation between attention and the predictive variable. That is, higher uncertainty leads to higher attention. This is because when uncertainty is close to zero, the investor observes the predictive coefficient and feels no incentive to pay attention and learn about it. The greater the uncertainty, the less accurate the investors’ estimates of the expected return, and therefore the larger their incentive to pay attention to news.

Panel (b) of Figure 1 shows that an increase in stock-return volatility decreases the curvature of the U-shaped relation between attention and the predictive variable. This is because an increase in the return volatility generates less informative returns, which decreases the volatility of expected returns (see Equation 21). Because in this case trends are more difficult to detect, the convexity of the value function in expected returns is less pronounced. Since the convexity of the value function determines the curvature of the U-shaped relation between attention and the predictive variable (see Equation 33), more return volatility leads to a weaker curvature and therefore to lower attention.\(^{16}\)

Panel (c) of Figure 1 shows that an increase in the absolute value of the predictive coefficient increases the curvature of the U-shaped relation between attention and the predictive variable. This is because large positive or negative values of the predictive coefficient imply a high expected return volatility (see Equation 21). Higher expected return volatility implies more opportunities to exploit trends, and thus a highly convex value function. This translates into a steeply curved U-shaped relation between attention and the predictive variable, and therefore into greater attention to news.

We now turn to the relation between attention and risk aversion, which is depicted in Figure 2. Consistent with Equation (33), an increase in risk aversion scales down the level of attention. This decreasing relation between attention and risk aversion comes from the

---

\(^{16}\)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 heads in the sand” when they expect to see bad news. Andries and Haddad (2014) provide an alternative explanation. In their model, investors are “disappointment averse” and thus are less attentive in riskier environments. In our case, riskier returns offer less marginal benefit for being attentive because the expected return becomes 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.
The figure depicts the relation between attention and the predictive variable for three different levels of risk aversion. State variables are $\hat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{V} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. Parameter values are provided in Table 1.

fact that an increase in risk aversion decreases the investor’s risky investment share (see the myopic component in Equation 32). The smaller the risky investment share, the lower the investor’s incentive to pay attention to news.

3.2.1 More on the impact of the predictive variable

We investigate the robustness of our results by performing a sensitivity analysis of the optimal attention with respect to changes in the dynamics of the predictive variable. More precisely, we analyze how attention responds to a change in the persistence and the volatility of the predictive variable.

Panels (a) and (b) of Figure 3 show that attention increases when both the persistence and the volatility of the predictive variable decrease. This is because the persistence and the volatility of the predictive variable determine the conditional volatility of its future values. When either the persistence or the volatility is low, the conditional volatility of future values of the predictive variable is low.\(^{17}\) Since the investor can efficiently exploit this “smoothness” to accurately predict future returns only if her estimate of the predictive coefficient is accurate, her optimal reaction to this situation is to pay more attention to news.

\(^{17}\)The conditional variance of future values of the predictive variable satisfies: $\text{Var}_{t,s}(y_s) = (1 - e^{-2\lambda_y(s-t)}) \sigma_y^2/2\lambda_y$. This function is increasing in $\sigma_y$ and decreasing in $\lambda_y$. 

18
Panels (a) and (b) depict the relation between attention and the predictive variable for three different levels of persistence of the predictive variable and its volatility, respectively. State variables are $\hat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. If not stated otherwise, parameter values are provided in Table 1.

### 3.3 Optimal risky investment share

Figure 4 plots the optimal risky investment share against the predictive variable for different values of uncertainty in panel (a), stock-return volatility in panel (b), and the estimated predictive coefficient in panel (c). The benchmark solid blue line is obtained by setting $\nu_t$, $V_t$, and $\hat{\beta}_t$ to their long-term levels $\nu_{ss}$, $\sqrt{\bar{V}}$, and $\bar{\beta}$.

Panel (a) of Figure 4 shows that the risky investment share is a hump-shaped function of the predictive variable and a decreasing function of uncertainty. This effect is driven by the hedging demand

$$ H_t^* \equiv w_t^* - \frac{\hat{\mu}_t - r^f}{\gamma V_t} = \frac{\nu_t(y_t - \bar{y})^2}{\gamma V_t} \frac{\phi_{\mu}}{\phi} < 0, \tag{36} $$

which reflects the investor’s willingness to hedge against variations in expected returns.\(^\text{18}\)

Since returns and expected returns co-move positively (see Equations 11 and 19), low returns imply low expected returns which triggers a negative hedging demand. Furthermore, the larger the expected return’s loading on return shocks, the more negative the hedging demand is. According to Equation (19), this loading decreases with both the predictive variable’s deviation from its mean and with uncertainty, and thus the hedging demand becomes more negative when the predictive variable’s deviation from its mean, and uncertainty, increase. This is confirmed by both Equation (36) and panel (a) of Figure 4. In addition, Equation

\(^\text{18}\)The hedging demand is negative because $J_{\mu} > 0$ and $J < 0$ imply that $\phi_{\mu} < 0$ and $\phi > 0$. 

---

**Figure 3: Impact of the dynamics of the predictive variable on attention.**

Panels (a) and (b) depict the relation between attention and the predictive variable for three different levels of persistence of the predictive variable and its volatility, respectively.

**Figure 4: Optimal risky investment share.**

Panels (a) and (b) depict the relation between attention and the predictive variable for three different levels of persistence of the predictive variable and its volatility, respectively. State variables are $\hat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. If not stated otherwise, parameter values are provided in Table 1.
Figure 4: Impact of the predictive variable on the risky investment share.
The relation between the risky investment share and the predictive variable. 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 $\beta_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{V} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. Parameter values are provided in
Table 1.

(36) shows that the hedging demand becomes less negative when the stock-return volatility
increases because the expected return’s loading on return shocks is decreasing with the
stock-return volatility.

Panel (b) of Figure 4 shows that the risky investment share decreases with stock-return
volatility (see myopic component in Equation 32). Furthermore, the risky investment share
increases with expected returns, which depend on the product of the predictive coefficient
and the demeaned predictive variable. That is, the risky investment share increases with
the predictive coefficient when the predictive variable is large and decreases with it when
the predictive variable is small. As panel (c) of Figure 4 shows, the risky investment share
is large whenever the product $\hat{\beta}_t(y_t - \bar{y})$ is positive and large.

3.4 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—
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 $\gamma$ and
information cost parameter $k$. The cost of ignoring news decreases with both risk aversion and the information cost parameter. A high risk aversion implies a small share invested in the stock and therefore a weak incentive to pay attention to learn about the stock’s expected return (see Figure 2). As a result, the optimal attention allocation strategy does not significantly differ from that of ignoring the news signal, i.e., the cost of ignoring information is small. Furthermore, the lower the cost parameter, the higher the optimal attention paid to news, and therefore the larger the cost of ignoring news. The cost of ignoring news can be significant, reaching as much as 5.1% of wealth when risk aversion and the information cost parameter are equal to 3 and 0.01, respectively.

<table>
<thead>
<tr>
<th>Cost parameter $k$</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk aversion $\gamma$ ↓</td>
<td>512.46</td>
<td>31.4</td>
<td>19.14</td>
</tr>
<tr>
<td>3</td>
<td>54.46</td>
<td>9.49</td>
<td>5.24</td>
</tr>
<tr>
<td>7</td>
<td>37.38</td>
<td>4.49</td>
<td>2.42</td>
</tr>
</tbody>
</table>

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 $\hat{\beta}_t = \bar{\beta} = 0$, $y_t = \bar{y} = 6.9\%$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \nu_{ss} = 2.11$. Parameter values are provided in Table 1.

4 Empirical analysis

In this section, we first show that there exists a positive and significant relation between the model-implied and empirical measures of attention and risky investment share. Furthermore, we provide evidence that attention significantly predicts the VIX. Then, we test the model’s predictions that attention is a U-shaped function of the predictive variable, an increasing function of both the absolute predictive coefficient and uncertainty, and a decreasing function of stock-return variance. We show that the data lend support to these theoretical predictions.

In Section 3.1, we used three state variables to calibrate the parameters of the model: the S&P 500 earnings-to-price ratio $y_t$, the time-varying predictive coefficient $\hat{\beta}_t$, and the conditional variance of returns $V_t$. Figure 5 illustrates the dynamics of these state variables, where the gray shaded areas represent NBER recessions.

In order to obtain model-implied time-series of attention and uncertainty, we first discretize the dynamics in Equation (16) as follows:

$$\nu_{t+\Delta} = \nu_t + \left[ - \left( \frac{(y_t - \bar{y})^2}{V_t} + a_t^* (\hat{\beta}_t, y_t, V_t, \nu_t) \right) \nu^2_t - 2\lambda_\beta \nu_t + \sigma^2_\beta \right] \Delta, \quad (37)$$
where $\Delta = 1/12 = 1$ month, the starting value $\nu_0$ is fixed to $\nu_{ss} = \sigma^2_{\beta}/(2\lambda_{\beta})$, and the optimal attention $a_t^*(\hat{\beta}_t, y_t, V_t, \nu_t)$ is a function of the state variables. Sequentially substituting the state variables depicted in Figure 5 in Equations (33), (37), and (32) yields: the model-implied uncertainty, the model-implied attention, and the model-implied risky investment share. These model-implied dynamics are depicted in the three panels of Figure 6.

We study the behavior of these model-implied quantities over the business cycle using NBER recession dummies. Table 3 shows that the model-implied attention is larger in recessions than in expansions. In recessions, the earnings-to-price ratio spikes and the predictive coefficient drops to negative values to reflect negative expected returns. Investors optimally
Figure 6: Historical dynamics of model-implied uncertainty, attention, and risky investment share.

Data are at a monthly frequency from 01/1955 to 12/2014.

react to these changes by paying more attention to news (as our theoretical model predicts in Figure 1). Attention is therefore counter-cyclical, consistent with Andrei and Hasler (2015). Table 3 also shows that investors place a smaller fraction of their wealth in the stock in recessions than in expansions. Finally, Table 3 shows that the model-implied uncertainty is pro-cyclical, in line with our theoretical prediction that higher attention tends to decrease uncertainty (as shown in Equation 16).

We then test whether the dynamics of model-implied attention, uncertainty, and risky investment share provide a realistic description of what is observed in the data. First, we build empirical proxies for these three measures. For the empirical measure of attention, \( a_t^E \), we select firms belonging to the Thomson-Reuters institutional database (13F) that
Table 3: Model-implied attention, risky investment share, and uncertainty in NBER recessions.

\(a_t^*, w_t^*, \nu_t\) are the model-implied measures of attention, risky investment share, and uncertainty, respectively. 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 a monthly frequency from 01/1955 to 12/2014.

<table>
<thead>
<tr>
<th></th>
<th>Attention (a_t^*)</th>
<th>Risky share (w_t^*)</th>
<th>Uncertainty (\nu_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.037*** (16.38)</td>
<td>0.288*** (5.00)</td>
<td>2.085*** (1208.20)</td>
</tr>
<tr>
<td>Recession dummy</td>
<td>0.019*** (3.30)</td>
<td>-0.255*** (-2.60)</td>
<td>-0.015*** (-2.76)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.029</td>
<td>0.009</td>
<td>0.028</td>
</tr>
<tr>
<td>Observations</td>
<td>720</td>
<td>720</td>
<td>720</td>
</tr>
</tbody>
</table>

For the empirical measure of uncertainty, \(\eta_t^E\), we use the 1-month-ahead macro-uncertainty index constructed by Jurado, Ludvigson, and Ng (2015). This time series is at a monthly frequency from 08/1960 to 12/2014. Note that this macro-uncertainty index is a measure of macroeconomic risk, whereas in our model \(\nu_t\) is the Bayesian uncertainty regarding the predictive coefficient \(\beta_t\), which measures learning inaccuracy rather than macroeconomic uncertainty. A proper measure of uncertainty, comparable with the Jurado et al. (2015) index, is the uncertainty about expected returns \(\eta_t\), which we define in Equation (20) of Lemma 1.

We proxy the empirical risky investment share, \(w_t^E\), by the negative of the log growth rate of the “Institutional Money Funds” index. The “Institutional Money Funds” index represents the dollar amount held by institutions in the money market and is obtained from

---

\(^{19}\)Empirical findings by Chordia and Swaminathan (2000), Lo and Wang (2000), Gervais, Kaniel, and Mingelgrin (2001), Barber and Odean (2008), and Fisher et al. (2016) suggest that trading volume is a reasonable proxy for investors’ attention.
the Federal Reserve Bank of St. Louis, at a monthly frequency from 10/1978 to 12/2014.

We then regress the empirical measures of attention, uncertainty, and risky investment share on their model-implied counterparts. Table 4 shows the results. There is a strong positive relation between the empirical and model-implied measures of attention, even after controlling for the autocorrelation in the empirical measure. This suggests that our model provides a realistic description of the dynamic attention allocation problem faced by investors. Furthermore, Table 4 shows a strong positive relation between the empirical and model-implied measures of uncertainty, but the relation loses significance after controlling for the autocorrelation in the empirical measure. As in Jurado et al. (2015), our model-implied uncertainty about expected returns is counter-cyclical. The correlation coefficient between the model-implied measure and the empirical measure is 0.6. Finally, Table 4 shows that the empirical and model-implied measures are strongly positively related, even after controlling for the autocorrelation in the empirical measure. This result suggests that, in addition to accurately describing investors’ dynamic attention behavior, our model also explains the dynamics of institutional investors’ risky investments.

We further test whether the model-implied and empirical measures of attention predict the VIX index. Table 5 shows that this is indeed the case. Both the model-implied and empirical measures of attention positively predict the VIX, even after controlling for the autocorrelation in the latter. That is, attention can be interpreted as a measure of future market risk. The positive relation between current attention and future VIX comes from the fact that attention is a proxy for the variance of expected returns, as we show in Equation (21). High attention today yields a high expected return variance, which indicates highly risky future returns. This is also consistent with recent findings by Fisher et al. (2016), who document that an increase in media attention positively relates to an increase in implied volatility.

### 4.1 Testing the predictions of the model

Our theoretical model predicts that investors’ attention is a U-shaped function of the predictive variable, a decreasing function of stock-return variance, and an increasing function of uncertainty. Furthermore, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand, which depends on uncertainty, the predictive variable, and the stock-return variance. In what follows, we provide empirical support for these theoretical predictions. We test the following two hypotheses.

---

---

20In unreported results, we also find a strong positive correlation (0.52) between the model-implied measure of uncertainty about expected returns and the VIX index.

25
Table 4: Empirical vs. model-implied attention, uncertainty and risky investment share.

The variables $a_t^E$, $\eta_t^E$, and $w_t^E$ denote the empirical measures of attention, uncertainty and risky investment share, respectively. The variables $a_t^*$, $\eta_t^*$, and $w_t^*$ denote the model-implied counterparts. We report Newey and West (1987) $t$-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***.

Uncertainty and attention data are at a monthly frequency from 08/1960 to 12/2014 and at the quarterly frequency from Q4/1983 to Q4/2014. The monthly model-implied attention $a_t^*$ is averaged over three consecutive months to obtain a quarterly measure. Risky investment share data are at a monthly frequency from 10/1978 to 12/2014.

Hypothesis 1. Equation (33) and the results depicted in Figure 1 show that attention can be approximated as follows:

$$a_t \approx C_0 + C_1 \nu_t^2 + \hat{C}_2(\hat{\beta}, V_t) \nu_t^2(y_t - \bar{y})^2 \approx C_0 + C_1 \nu_t^2 + C_2 \frac{|\hat{\beta}_t|}{V_t} \nu_t^2(y_t - \bar{y})^2,$$  

(38)

where $C_1 < 0$ and $C_2 > 0$. That is, the curvature of the U-shaped relation between attention and the demeaned predictive variable decreases with the stock-return variance and increases with both the absolute predictive coefficient and the squared uncertainty. In addition, the squared uncertainty determines the level of the U-shaped relation.

Hypothesis 2. Equation (32) and the results depicted in Figure 4 show that the risky in-
Table 5: The predictive power of attention on the VIX.

<table>
<thead>
<tr>
<th></th>
<th>VIX&lt;sub&gt;t&lt;/sub&gt;</th>
<th>VIX&lt;sub&gt;t&lt;/sub&gt;</th>
<th>VIX&lt;sub&gt;t&lt;/sub&gt;</th>
<th>VIX&lt;sub&gt;t&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.153***</td>
<td>0.029***</td>
<td>0.196***</td>
<td>0.059***</td>
</tr>
<tr>
<td></td>
<td>(12.70)</td>
<td>(5.06)</td>
<td>(18.48)</td>
<td>(7.31)</td>
</tr>
<tr>
<td>a&lt;sub&gt;t−1/12&lt;/sub&gt;</td>
<td>1.123***</td>
<td>0.148*</td>
<td>0.700***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.59)</td>
<td>(1.95)</td>
<td>(20.77)</td>
<td></td>
</tr>
<tr>
<td>VIX&lt;sub&gt;t−1/12&lt;/sub&gt;</td>
<td>0.823***</td>
<td>0.040***</td>
<td></td>
<td>0.008**</td>
</tr>
<tr>
<td></td>
<td>(20.03)</td>
<td>(7.42)</td>
<td></td>
<td>(2.24)</td>
</tr>
<tr>
<td>a&lt;sub&gt;E−1/12&lt;/sub&gt;</td>
<td>0.823***</td>
<td>0.040***</td>
<td></td>
<td>0.008**</td>
</tr>
<tr>
<td></td>
<td>(20.03)</td>
<td>(7.42)</td>
<td></td>
<td>(2.24)</td>
</tr>
<tr>
<td>R&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.201</td>
<td>0.729</td>
<td>0.107</td>
<td>0.530</td>
</tr>
<tr>
<td>Observations</td>
<td>299</td>
<td>299</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

The variables a<sub>E</sub> and a<sub>t</sub> denote the empirical and model-implied measures of attention, respectively. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. Model-implied and empirical attention data are at the monthly frequency from 01/1990 to 12/2014 and at a quarterly frequency from Q1/1990 to Q4/2014.

The risky investment share can be approximated as follows:

\[
w_t \approx K_0 + K_1 \frac{\bar{\mu} + \hat{\beta}_t(y_t - \bar{y}) - r^f}{V_t} + K_2 \frac{\nu_t(y_t - \bar{y})^2}{V_t}, \tag{39}
\]

where \( K_1 > 0 \) and \( K_2 < 0 \). That is, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand that is hump-shaped in the demeaned predictive variable. The curvature of the hedging demand decreases with the stock-return variance and increases with uncertainty.

The parameters \( C_0, C_1, C_2, K_0, K_1, \) and \( K_2 \) are estimated using ordinary least squares. The left-hand side of Equations (38) and (39) define the dependent variables, while the right-hand side define the independent variables.

We report results regarding Hypothesis 1 in Table 6. The second and third columns test Hypothesis 1 with the model-implied values of attention. The coefficients \( C_1 \) and \( C_2 \) are indeed negative and positive, respectively. Comparing the \( R^2 \) obtained in these two columns shows that adjusting the level of the U-shaped relation between attention and the demeaned predictive variable with the squared uncertainty only has a marginal impact on the goodness of fit. The fourth and fifth columns show the coefficients obtained using the empirical measure of attention. The numbers lend support to the model-implied relation between attention and the state variables, whose curvature decreases with the stock-return variance and increases with both the absolute value of the predictive coefficient and the squared uncertainty. In addition, the coefficient \( C_1 \) is negative and statistically insignifi-
Table 6: Model-implied and empirical measures of attention vs. predictive variable, predictive coefficient, return variance, and uncertainty.

The variables $a^E_t$ and $a^*_{t}$ are the empirical and model-implied measures of attention, respectively. We report Newey and West (1987) $t$-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***.

Model-implied and empirical attention data are at a monthly frequency from 10/1983 to 12/2014 and at a quarterly frequency from Q4/1983 to Q4/2014, respectively. In the fourth and fifth columns, the monthly model-implied predictive variable $y_t$, predictive coefficient $\hat{\beta}_t$, return variance $V_t$, and uncertainty $\nu_t$ are averaged over three consecutive months to obtain quarterly measures.

Recent empirical evidence by Fisher et al. (2016) supports the U-shaped relation predicted by our model. Fisher et al. (2016) build indices of media attention on various macroeconomic fundamentals and show that attention varies with movements in these fundamentals. Importantly, when testing for a quadratic relation between attention and detrended moving averages of macroeconomic fundamentals, Fisher et al. (2016) find that squared terms are important and significantly positive, consistent with our idea that attention rises when economic fundamentals depart from their average values.

We report results regarding Hypothesis 2 in Table 7. The second and third columns test Hypothesis 2 with the model-implied risky investment share. The coefficients $K_1$ and $K_2$ are indeed positive and negative, respectively. The third column shows that the model-implied risky investment share increases with the Sharpe ratio of the stock and features a negative but statistically insignificant hedging demand. The fourth and fifth columns of Table 7 lend support to the model’s prediction. Indeed, the empirical measure of risky investment share increases with the Sharpe ratio of the stock and features a negative and significant hedging demand. As in the model, the curvature of the hump-shaped relation between the empirical hedging demand and the demeaned predictive variable decreases with the stock-
Risky share \( w_t^* \) and \( w_t^E \) are the empirical and model-implied measures of risky investment share, respectively. We report Newey and West (1987) \( t \)-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***.

Data are at a monthly frequency from 10/1978 to 12/2014.

Table 7: Model-implied and empirical risky investment share vs. Sharpe ratio and hedging demand.

The variables \( w_t^E \) and \( w_t^* \) are the empirical and model-implied measures of risky investment share, respectively. We report Newey and West (1987) \( t \)-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***.

Data are at a monthly frequency from 10/1978 to 12/2014.

The model-implied and empirical risky investment share vs. Sharpe ratio and hedging demand.

\[
\begin{array}{|c|c|c|c|c|}
\hline
& \text{Risky share } w_t^* & \text{Risky share } w_t^E & \text{Risky share } w_t^E & \text{Risky share } w_t^E \\
\hline
\text{Intercept} & 0.001 & 0.002 & -0.017 & -0.007 \\
& (0.13) & (0.09) & (-7.53) & (-3.91) \\
\frac{\hat{\beta}_t(y_t - \bar{y}) - r^f}{V_t} & 0.337*** & 0.337*** & 0.003*** & 0.002*** \\
& (48.51) & (45.68) & (2.83) & (3.09) \\
\frac{\nu_t(y_t - \bar{y})^2}{V_t} & -0.002 & (-0.01) & -0.083*** & (-5.39) \\
\hline
R^2 & 0.970 & 0.970 & 0.044 & 0.217 \\
Observations & 435 & 435 & 435 & 435 \\
\hline
\end{array}
\]

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
Table 8: Model-implied and empirical relation between risky investment share and attention.

<table>
<thead>
<tr>
<th></th>
<th>Risky share $w_t^*$</th>
<th>Risky share $w_t^{FE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.464***</td>
<td>−0.008***</td>
</tr>
<tr>
<td></td>
<td>(3.38)</td>
<td>(−2.95)</td>
</tr>
<tr>
<td>$\beta_t(y_t - \bar{y})a_t^*$</td>
<td>198.50***</td>
<td>0.175***</td>
</tr>
<tr>
<td></td>
<td>(3.94)</td>
<td>(4.41)</td>
</tr>
<tr>
<td>$\beta_t(y_t - \bar{y})a_t^{FE}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.488</td>
<td>0.075</td>
</tr>
<tr>
<td>Observations</td>
<td>375</td>
<td>125</td>
</tr>
</tbody>
</table>

The variables $a_t^F$ and $a_t^*$ are the empirical and model-implied measures of attention, respectively. The variables $w_t^F$ and $w_t^*$ are the empirical and model-implied measures of risky investment share, respectively. We report Newey and West (1987) $t$-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***.

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, a decreasing function of stock-return volatility, and an increasing function of the absolute predictive coefficient. In addition, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand. The hedging demand is a hump-shaped function of the return predictor, an increasing function of stock-return volatility, and a decreasing function of uncertainty. We test these theoretical predictions and show that the data lend support to most of them.

Our analysis can be extended to a multiple asset setting, which would help understand the impact of costly dynamic information acquisition on diversification. It would be also interesting to investigate the impact of the optimal choice of attention on the equilibrium risk-free rate, equity premium, and equity return volatility in a pure-exchange economy. Finally, a production economy can offer insights on the impact of costly information acquisition on the dynamics of aggregate consumption.
References


A Appendix

A.1 Proof of Proposition 2

The dynamics of the vector of state variables \( Z_t = [\hat{\mu}_t, y_t, V_t, \nu_t]^\top \) satisfy

\[
dZ_t = m_t dt + \Sigma_{1,t} d\hat{B}^1_{1,t} + \Sigma_{2,t} \begin{pmatrix} d\hat{B}^2_{2,t} \\ d\hat{B}^3_{3,t} \\ d\hat{B}^4_{4,t} \end{pmatrix},
\]

where the 4-dimensional vector of drift \( m \), the 4-dimensional vector of diffusion \( \Sigma_1 \), and the \( 4 \times 3 \) matrix of diffusion \( \Sigma_2 \) satisfy

\[
m = \begin{pmatrix} (\lambda_y + \lambda_\beta) \left( \bar{\mu} + \frac{\beta \lambda_\beta (y-\bar{y})}{\lambda_y + \lambda_\beta} - \hat{\mu} \right) \\
\lambda_y (\bar{y} - y) \\
\lambda_V (\bar{V} - V) \\
- \left( \frac{(y-\bar{y})^2}{V} + a \right) \nu^2 - 2 \lambda_\beta \nu \end{pmatrix}
\]

\[
\Sigma_1 = \begin{pmatrix} \frac{(y-\bar{y})^2 \nu}{\sqrt{V}} \\
0 \\
0 \\
0 \end{pmatrix}
\]

\[
\Sigma_2 = \begin{pmatrix} \sigma_y \frac{\bar{\mu} - \bar{\mu}}{\bar{y} - \bar{y}} & 0 & \nu \sqrt{a (y - \bar{y})} \\
\frac{\sigma_y}{\bar{y} - \bar{y}} & 0 & 0 \\
0 & \sigma_V \sqrt{V} & 0 \\
0 & 0 & 0 \end{pmatrix}
\]

Using this notation, we can write the optimality condition for (23) as

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

where \( \mathcal{D}^{W,Z} \) is the infinitesimal generator such that\(^{21}\)

\[
\mathcal{D}^{W,Z} J = J'_2 m + J'_W \left( (r^f - K(a_t))W_t + w_t W_t (\bar{\mu}_t - r^f) - c_t \right) + \frac{1}{2} J_W W_t w_t V_t + \frac{1}{2} \text{tr} \left[ (\Sigma_1 \Sigma_1' + \Sigma_2 \Sigma_2') J_{ZZ} \right] + W_t w_t \sqrt{V_t} \Sigma_1' J_{WZ}.
\]

Differentiating (23) partially with respect to the control variables yields the first order conditions:

\[
0 = u_c - J_W
\]

\[
0 = J_W W_t (\bar{\mu}_t - r^f) + J_{WW} W_t^2 W_t w_t + J_{WV} W_t \nu_t (y_t - \bar{y})
\]

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

Solving the first order conditions yields Proposition 2.

\(^{21}\)Note that, for notational convenience, we drop hats when state variables appear as indices.
A.2 Numerical solution method

Substituting the first-order conditions in Equation (43) yields a PDE the function $\phi(\mu, y, V, \nu)$ has to satisfy. We numerically solve this PDE using the Chebyshev collocation method described in Judd (1998). That is, we approximate the function $\phi(\mu, y, V, \nu)$ as follows:

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

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 $P(\mu, y, V, \nu)$ 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.