Limited Attention and the Dynamics of Probability Weighting*

Nikolaus Schweizer       Martijn de Vries

January 10, 2020

Preliminary and incomplete – please do not circulate

Abstract

We study the hypothesis that, due to frictions like cognitive limits, investors are selective in their attention allocation. Using survey data we find that investors attention allocation is based on how remarkable a return realization is compared to its recent past. Based upon our empirical findings, we propose a simple theoretical model of inattentive learning. We derive the limiting distortion, the subjective mean and the subjective variance in closed form. Our limiting probability distortion functions form a new class of highly tractable polynomial probability weighting functions with interpretable parameters. Furthermore, we find that inattentive learning generates several familiar belief distortions like probability weighting as in prospect theory, a preference for positive skewness, overestimation of volatility and overextrapolation of returns. We perform cross-sectional tests on individual stocks and find significant pricing effects.

Key words: Limited Attention, Probability Weighting, Inattentive Learning, Distorted Beliefs, Overextrapolation

JEL classification: D91, G41

* Nikolaus Schweizer is at the department of Econometrics and Operations Research at Tilburg University. Martijn de Vries is at the Department of Finance at Tilburg University. For helpful comments and suggestions we want to thank Nicholas Barberis, Stefano Cassella, Emiel Jerphanion, Lawrence Jin, Frank de Jong, Jens Kvaerner, Judith C. Schneider, Bas Werker and Ole Wilms. Please send correspondence to: m.a.devries_1@tilburguniversity.edu.
1 Introduction

It is common for models in financial economics to assume rational expectations and Bayesian updating. However, it has been shown in many occasions that investors on financial markets have biased expectations (Greenwood & Shleifer, 2014; Bordalo et al., 2017; Barberis et al., 2018; Cassella & Gulen, 2018). Examples of such biases are, probability weighting, overconfidence, loss aversion, overextrapolative beliefs, and many others. Yet little is known where many of these biases originate from.\footnote{For example Gervais, Kaniel, and Mingelgrin (2001) show that self-attribution is a driver of overconfidence and Barberis (2013) argues that the representativeness heuristic is able to generate overextrapolative beliefs.}

The contribution of our paper to this literature is fourfold. First, we test the empirical relevance of limited attention models by verifying their implications. Second, we propose a model of inattentive learning which we base on our empirical verification. Third, we derive the implications of the model of inattentive learning by means of closed-form expressions of the limiting belief distortion, the subjective mean and subjective variance. We find that inattentive learning is able to generate multiple behavioral biases simultaneously, including probability weighting, preference for skewness and overextrapolative belief formation. Fourth, we test on financial markets using individual stocks for pricing implications. We find that stocks that a more favorable become overpriced and therefore earn subsequently lower returns. This effect is strong for daily horizons and to a lower extent still present for monthly horizons.

We show the empirical relevance of limited attention by testing its predictions using survey data. We are interested in belief updating and therefore look at changes in the beliefs reported in the surveys. To better understand the learning dynamics over time, we direct our interest to the relation between information and changes in beliefs. We argue that investor beliefs of future returns of the financial market as measured by survey data provides a useful setting to test for limited attention. The hypothesis that we test is that investors direct more attention to extreme return realizations. Under the assumption that financial markets are efficient, all information should be reflected by prices. Additionally, as returns are widely available, one may expect investor expectations to be largely influenced by historical returns. In our analysis we therefore quantify the relation of lagged returns on changes in expectations as measured by survey data. When sorting the lagged returns, we find that only the extreme returns are significantly effecting belief changes, which confirms our hypothesis.

Taking our empirical findings to theory, we propose a simple theoretical model. The model works as follows. When a new returns realizes the investor attaches an attention weight to the return based on the recent past returns. We propose the investor uses an evaluation window
of a given length and the new return realization is ranked among the returns in this window. The attention weight the new return obtains depends on its rank relative to the evaluation window. This learning dynamics is able to incorporate that investors direct more attention to return realizations that are extreme relative to recent returns. Such behavior is captured by having an attention weighting function that attaches relatively high weights to the higher and the lower ranks relative to the ranks in the middle. When assuming investors only direct more attention weight to the largest and smallest ranks, i.e. the maximum and minimum, the model simplifies to a three-parameter specification.

Our model of learning under limited attention generates biased beliefs that persist even after observing an infinite number of observations. We derive the limiting distortion, the subjective mean and the subjective variance in closed-form. The limiting distortion is very much related to the static probability weighting function proposed by Kahneman and Tversky (1979). Our probability weighting distortion is easily linked to the parameters of the attention weighting function. For specific parameters we are able to closely match the functional form of the Kahneman and Tversky (1979) probability weighting function. In addition to connecting limited attention to probability weighting, the distortion of the subjective mean reveals the emergence of a preference for skewness. The closed-form solution to the subjective variance highlights that for reasonable parameters, investors will overestimate variance. Lastly, our model naturally generates overextrapolation. When a new return of a relatively large magnitude realizes, an unlimited rational investor that using Bayesian updating will shift his beliefs upward, which causes a minor extrapolative effect. However, in case of learning under limited attention the investor will attach significantly more weight to this return and therefore overextrapolate in either the positive or negative direction.

In this paper we use survey data to better understand the belief dynamics of investors over time. Survey data has been shown to strongly correlate with aggregate investors expectations (Greenwood & Shleifer, 2014). Due to these findings we are confident that analyzing survey data will provide a better picture of what is driving aggregate expectations. The task of determining the driving force behind investor expectations is important as it may provide an explanation for why the behavioral assumptions noted above are generated. The necessity of unifying behavioral biases stems from the lack of discipline critique that has been widely used against the behavioral finance literature over the past (Barberis & Thaler, 2003). It claims that for any anomaly in financial markets one is able to flip a psychology book and find a bias that maps one-to-one to this anomaly. Additionally, a widely used argument against the biases put forth by the behavioral literature is that it is unlikely that such biases are present.
once beliefs are aggregated. Our analysis will be helpful to address these concerns as we make use of aggregate beliefs.

When one thinks of investors acting in financial markets, it is noticeable how many stocks are available to them. Even without thinking about all the tasks an investor, and retail investors more strongly, have to perform during any day, the action on the market and all information produced can be quite exhausting. It is therefore plausible that investors do not pay attention to all stocks equally at each point in time. Bordalo et al. (2012) show that investors will cross-sectionally focus on stocks based on their salience, which is high when a stock performs well when all other stocks perform poorly. The underlying motivation of salience, that investors are selective in allocating their attention, is similar to our focus. However, in our paper we focus on the dynamics of beliefs over time instead of the cross-sectional implications.

We are not the first to stress the importance of limits to attention (Sims, 2006). There exist models where investors are assumed to exhibit limited attention, or other frictions, such that Bayesian updating will be costly (Kominers et al., 2018). Intuitively, the overall prediction of this literature is that under limited attention, agents will direct their attention towards observations that are more remarkable compared to their current beliefs and/or recent observations. The reasoning behind these predictions is as follows. Take the example of costly Bayesian updating and assume a setting of an investors in the equity market who adapts his beliefs of the underlying distribution when new returns realize. Assume that whenever the investors updates his beliefs he pays a fixed cost, which for example represents the cost of actively paying attention. Subconsciously the investor perceives all return realizations, though, he only actively updates his beliefs in case the increase of his expected utility outweigh the costs. It is straightforward to notice that he will therefore only actively update his beliefs in case the return realization is at least some distance away of his current beliefs. The investor will therefore update his beliefs for the relative extreme realizations.

The paper is organized as follows. Section 2 analyzes survey data to verify the implications put forth by the theoretical literature on limited attention. Section 3 proposes a model of learning under limited attention based on the empirical findings of Section 2. Furthermore, Section 3 shows the theoretical implications of the model. In section 4 we test for empirical relevance in the stock market using individual stocks and showing presence of the implications of inattentive learning. Section 5 concludes.
2 Attention allocation

The assumption that we make throughout this paper is that attention is limited. Subsequently, we ask ourselves the question how this affects belief formation. Models of limited attention suggest that individuals direct their attention towards observations that are further away from their prior, i.e., towards extreme realizations (Kominers et al., 2018). In this section we empirically test this prediction. We use investors expectations and determine how these are influenced by historical returns. To measure beliefs we follow Greenwood and Shleifer (2014) and rely on survey data. The historical returns will be market returns and we are interested in how investors update their beliefs depending on recent past returns.

The analysis in this section is twofold. First, we verify the prediction that investors have limited attention and because of that direct their attention to extreme return realizations. These results provide empirical relevance of modeling limited attention and are therefore of main importance for our paper. Second, we determine how survey beliefs are affected by lagged return realizations and more specifically how far the memory of the investors stretches. For example, Barberis et al. (2016) use monthly data and use a memory window of 60 months. This is proposed based on the five years of historical data that investors commonly get presented in their brokerage accounts. In our setting, we use daily returns and we measure the memory length of investors using the survey data. We hypothesize that for daily decision making, investors have an active memory length that is relatively short compared to the lower-frequency decision making in Barberis et al. (2016).

2.1 Verification from Survey data

The data we use to measure beliefs is a survey collected by the American Association of Individual Investors (AAII). The survey represents expectations of retail investors on a weekly basis from the 24th of July 1987 up to 14th of March 2019. Each week investors may file their answer to the question whether they expect the stock market to go up, down or stay the same over the next six months. The investors are then classified as bullish, bearish or neutral, respectively. From the weekly pool of investors, a random sample of investors is drawn and the final available data represents the aggregate percentage of investors of this sample that is bullish, bearish or neutral.²

Survey data allows us to determine a link between the information given to investors and

²Data can be downloaded from www.aaii.com. For more details on the survey data see Greenwood and Shleifer (2014).
their beliefs. Namely, under efficient capital markets prices should reflect all information, which makes historical returns a good measure of information. Also, returns on the stock market are widely available to investors and thus one may expect returns to be an important source for the belief formation of future returns. We therefore use the returns of the S&P500 index as a measure of information.

The AAII-survey data is measured on a weekly frequency. We investigate how lagged returns affect changes in belief formation as measured by either bullish or bearish.\(^3\) In between two survey dates investors observe five daily return realizations. In case the survey participants would be a closed panel of which all participants file their answer each week, the weekly changes in survey outcomes should, in principle, only be influenced by the intermediate return realizations. Unfortunately, the data does not show at which day each investor files his survey and how many investors are filing the weekly survey repeatedly. In case few investors file subsequent surveys, returns before the previous survey may also affect the weekly change simply because these returns affected the beliefs of investors that were not present in the previous survey. In the analysis we will therefore include more than the recent five daily lagged returns.\(^4\)

As we are interested in the belief updating, we use weekly belief changes which we compute as follows,

\[
\Delta \text{bullish}_t = \frac{\text{bullish}_t - \text{bullish}_{t-1}}{\text{bullish}_{t-1}}, \quad \Delta \text{bearish}_t = \frac{\text{bullish}_t - \text{bullish}_{t-1}}{\text{bullish}_{t-1}}
\]  

(1)

Subsequently, we regress these weekly belief changes on daily lagged returns. As we expect the group of investors underlying the belief changes to vary significantly from one week to another, we include up to 10 lagged returns. The results are in Table 1.

The first and second regression in Table 1 show our baseline verification results. Notable is the insignificance of the first lagged return. This is easily explained by the possibility of investors to file the survey any time during the week. As there is no restriction on when to file the survey, it seems likely that investors file their survey before the end of the week. The results thus show that few investors filed the survey after the markets were closed at the final day. We do observe that lags two up to seven are of significant importance for changes in beliefs. The significance of sixth and seventh lagged return is due to the change

\(^3\)In our analysis we do not include the neutral category. It seems intuitively unlikely that an investor expects the future return to be exactly zero. Therefore, there will be bounds between which investors choose the no-change option. As the other two groups are more cleanly identified we restrict our analysis to them.

\(^4\)Furthermore, in Appendix C we show results for bi-weekly changes. These are computed similar to 1 when using two weeks. The results are presented in Table 7, which are similar to those in Table 1.
<table>
<thead>
<tr>
<th></th>
<th>$r_{t-1}$</th>
<th>$r_{t-2}$</th>
<th>$r_{t-3}$</th>
<th>$r_{t-4}$</th>
<th>$r_{t-5}$</th>
<th>$r_{t-6}$</th>
<th>$r_{t-7}$</th>
<th>$r_{t-8}$</th>
<th>$r_{t-9}$</th>
<th>$r_{t-10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta$Bullish</td>
<td>$\Delta$Bearish</td>
<td>$\Delta$Bullish</td>
<td>$\Delta$Bearish</td>
<td>$\Delta$Bullish</td>
<td>$\Delta$Bearish</td>
<td>$\Delta$Bullish</td>
<td>$\Delta$Bearish</td>
<td>$\Delta$Bullish</td>
<td>$\Delta$Bearish</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>-0.41</td>
<td>0.30</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(-0.48)</td>
<td>(0.49)</td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.46***</td>
<td>-3.19***</td>
<td>2.29***</td>
<td>-2.75***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.62)</td>
<td>(-4.04)</td>
<td>(4.19)</td>
<td>(-3.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.70***</td>
<td>-3.68***</td>
<td>2.51***</td>
<td>-3.22***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.71)</td>
<td>(-5.20)</td>
<td>(5.24)</td>
<td>(-4.68)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.99***</td>
<td>-4.90***</td>
<td>2.84***</td>
<td>-4.52***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.13)</td>
<td>(-4.95)</td>
<td>(3.92)</td>
<td>(-4.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.08***</td>
<td>-2.18***</td>
<td>1.91***</td>
<td>-1.75**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.09)</td>
<td>(-2.78)</td>
<td>(3.54)</td>
<td>(-2.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.75***</td>
<td>-3.78***</td>
<td>2.60***</td>
<td>-3.40***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.39)</td>
<td>(-5.43)</td>
<td>(4.88)</td>
<td>(-4.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.12***</td>
<td>-2.26***</td>
<td>1.94***</td>
<td>-1.79**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.20)</td>
<td>(-3.30)</td>
<td>(3.74)</td>
<td>(-2.55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>-1.49*</td>
<td>0.11</td>
<td>-1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td>(-1.92)</td>
<td>(0.20)</td>
<td>(-1.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.28</td>
<td>0.60</td>
<td>-0.42</td>
<td>0.98*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.53)</td>
<td>(1.06)</td>
<td>(-0.82)</td>
<td>(1.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.12**</td>
<td>-0.62</td>
<td>0.97*</td>
<td>-0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.03)</td>
<td>(-0.87)</td>
<td>(1.73)</td>
<td>(-0.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02***</td>
<td>0.05***</td>
<td>0.00</td>
<td>0.08***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.34)</td>
<td>(9.06)</td>
<td>(0.46)</td>
<td>(6.38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pos-dum</td>
<td>-</td>
<td>-</td>
<td>0.02</td>
<td>-0.05***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.65)</td>
<td>(-2.69)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>1647</td>
<td>1647</td>
<td>1647</td>
<td>1647</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>adj. $R^2$</td>
<td>0.0740</td>
<td>0.0654</td>
<td>0.0747</td>
<td>0.0687</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The table reports the results from the regression of changes in beliefs on lagged returns. Changes in beliefs are computed as given by (1) using weekly survey data and lagged daily returns of the S&P500. The regression therefore takes into account two weeks of lagged returns. Newey and West (1987)-standard errors are reported in brackets.

of composition of investors filing the survey in two subsequent weeks. This pattern of the significance of lagged returns is the same for both beliefs measures. The magnitude of the regression coefficients is not our main interest, though, we do observe that the signs of the coefficients are in line with our intuition. For the beliefs measured by bullish all significant coefficients are positive and for the bearish measure all significant coefficients are negative. These signs imply that higher lagged returns increase (decrease) the change of the proportion of investors being bullish (bearish). These results basically show that after larger returns investors adjust their expectation for future returns upwards and for smaller returns they
adjust their expectations downwards.

In the third and fourth regression we add the pos-dum variable. This is a dummy variable that equals one if there are more positive than negative returns over all lags taken into account in the regression. We introduce this variable due to the specific characterization of the AAII-survey. Namely, the survey asks a qualitative question on whether the stock market will go either up or down. We thus include this dummy variable to address the concern that some investors might simply count the number of lagged returns with either sign to form expectations. We will refer to this particular way of answering the survey as the median-effect.

We hypothesized that investors are subject to limited attention and because of that they allocate attention towards extreme return realizations. To test this prediction more precisely we proceed our analysis by accounting for the magnitude of lagged returns. We use the lagged returns that significantly impact belief changes as documented in Table 1 and we sort them within each weekly observations.\footnote{The desctiptive statistics of the sorted lagged returns are in Table 6 in Appendix B.} Similar to the approach of Table 1 we then regress the belief changes on the sorted lagged returns. The results of these regressions are in Table 2.

The regression results in Table 2 show that relatively extreme lagged returns have a very significant impact on belief changes of the investor. The results are most clearly shown for belief changes as measured by bullishness. From the second regression it appears that the median-effect is more expressed for the beliefs as measured by bearishness. However, when controlling for this effect with our pos-dum variable, we observe that the results using both bullishness and bearishness are in line with our hypothesis. To show further robustness of these results, we run similar regressions when adjusting the number of lags used in the regression. All results are similar in the sense that the extreme lagged returns significantly impact changes in beliefs. These results are shown in Tables 9 and 10 in Appendix C.

Our results are in line with our hypothesis that investors direct attention based on the magnitude of a realization. We argue that this behavior may originate from cognitive limits. We cannot rule out alternative causes that for example investors allocate their attention to extreme returns because they expect such realizations to convey relatively more information. Our contribution is not to determine the sole cause of this behavior, instead we are interested in characterizing the dynamics and derive it’s implications. In the next section we therefore propose a belief dynamics that incorporates our empirical findings.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th></th>
<th>(2)</th>
<th></th>
<th>(3)</th>
<th></th>
<th>(4)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔBullish</td>
<td>ΔBearish</td>
<td>ΔBullish</td>
<td>ΔBearish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{1:6}$</td>
<td>3.21***</td>
<td>-3.01**</td>
<td>3.06***</td>
<td>-2.87**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.48)</td>
<td>(-2.51)</td>
<td>(4.25)</td>
<td>(-2.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{2:6}$</td>
<td>4.08***</td>
<td>-4.94**</td>
<td>4.01***</td>
<td>-4.87**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.09)</td>
<td>(-2.29)</td>
<td>(3.07)</td>
<td>(-2.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{3:6}$</td>
<td>1.23</td>
<td>-1.39</td>
<td>1.08</td>
<td>-1.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(-0.67)</td>
<td>(0.61)</td>
<td>(0.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{4:6}$</td>
<td>2.38</td>
<td>-4.51**</td>
<td>0.75</td>
<td>-3.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(-2.21)</td>
<td>(0.45)</td>
<td>(1.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{5:6}$</td>
<td>1.99</td>
<td>-4.12**</td>
<td>1.74</td>
<td>-3.89*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.59)</td>
<td>(-2.07)</td>
<td>(1.37)</td>
<td>(-1.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{6:6}$</td>
<td>2.29***</td>
<td>-2.09*</td>
<td>2.32***</td>
<td>-2.13**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
<td>(-1.91)</td>
<td>(3.21)</td>
<td>(-1.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.00</td>
<td>0.06***</td>
<td>-0.02**</td>
<td>0.07***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.72)</td>
<td>(6.63)</td>
<td>(-2.10)</td>
<td>(5.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pos-dum</td>
<td>-</td>
<td>-</td>
<td>0.03**</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.21)</td>
<td>(-1.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1647</td>
<td>1647</td>
<td>1647</td>
<td>1647</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.0764</td>
<td>0.0598</td>
<td>0.0786</td>
<td>0.0603</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The table reports the results from the regression of changes in beliefs on sorted lagged returns, where the lags used are of two up to seven days. Changes in beliefs are as in Equation 1 using weekly survey data and lagged returns are daily returns on the S&P500. $r_{1:6}$ represents the maximum lagged return and $r_{6:6}$ the minimum. Newey and West (1987)-standard errors are reported in brackets.

3 A Model of Inattentive Learning

3.1 General Model

Let us now introduce in detail our tractable mathematical model of belief formation under rank-dependent attention. There is a sequence of observations given by real-valued random variables $X = (X_j)_{j=1}^n$ each drawn from a cumulative distribution function $F$. Unless otherwise noted, we also assume that the $X_i$ are statistically independent and drawn from a continuous distribution. An agent observes the sequence sequentially, first $X_1$, then $X_2$, $X_3$ and so on. The agent wishes to learn the distribution of the $X_i$. To this end, the agent evaluates each new, incoming observation and updates his belief about the distribution. During each evaluation, the agent focuses on the latest $n$ observations in the sequence, where we call $n$

---

6 Due to the continuous nature of the distribution, we do not have to worry about tie breaking that would occur when computing ranks in a sample with repeated observations. To complete the model, we assume that any ties that occur are solved randomly.
the evaluation horizon of the agent, \( m > n \). Subsequent to ranking all \( n \) realizations in the evaluation horizon, the most recent realization obtains an attention weight according to \( \omega \), which we refer to as the rank-dependent attention weight function,

\[
\omega : \{1, \ldots, n\} \to [0, 1], \quad \text{s.t.} \quad \sum_{i=1}^{n} \omega(i) = 1
\]  

(2)

This attention allocation rule is applied dynamically as each new realization is ranked among the \( n - 1 \) previous realizations and obtains a weight according to its rank. Note that the first evaluation attaches weights to the first \( n \) realizations and in subsequent periods only the new realization obtains a new weight and the others only change because of normalization. Each observation is forever associated with the attention it got when it arrived. In this sense, the agent’s learning process is governed by limited attention. Define the ranking function \( \rho_n(k, x) \) via

\[
\rho_n(k, x) = \begin{cases} 
\text{Rank of } x_k \text{ among } x_{k-n+1}, x_{k-n+2}, \ldots, x_k \text{ if } k \geq n \\
\text{Rank of } x_k \text{ among } x_1, \ldots, x_k \text{ if } k < n
\end{cases}
\]

(3)

Then, after \( m \) observations, an agent who uses our updating rule and observes the realization \( x \) believes that the empirical distribution function is equal to

\[
\hat{F}_{\omega,m}(c|x) = \frac{\sum_{k=1}^{m} \omega(\rho_n(k, x)) \mathbb{1}_{x_k \leq c}}{\sum_{k=1}^{m} \omega(\rho_n(k, x))}
\]

(4)

where the indicator \( \mathbb{1}_{x_k \leq c} \) equals one in case \( x_k \leq c \) and zero otherwise. Thus, \( \hat{F}_{\omega,m}(c|x) \) is the agent’s estimate of the probability that the next observation will be at most \( c \). As our learning rule is based on an evaluation window of length \( n \), we have to define the rule slightly differently for the first \( n \) observations. However, for most of our results, such as the limit analysis of Proposition 1, the ranking rule for the first \( n \) observations is irrelevant.

For the sake of analytic convenience, we will also consider the alternative distorted empirical distribution function

\[
\tilde{F}_{\omega,m}(c|x) = \frac{n}{m} \sum_{k=1}^{m} \omega(\rho_n(k, x)) \mathbb{1}_{x_k \leq c}
\]

(5)

which differs from (4) as the agent is not normalizing after every new observation is observed, using instead an average weight of \( \frac{n}{m} \). Mathematically, the belief distribution \( \tilde{F}_{\omega,m}(c|x) \) is slightly easier to analyze than \( \hat{F}_{\omega,m}(c|x) \). However, unlike \( \hat{F}_{\omega,m}(c|x) \), it can take values outside the interval \([0, 1]\) and is thus not a proper probability distribution. Proposition 1 shows
however that asymptotically, as the number of observations \( m \) grows, both distorted belief distributions converge to a probability-weighted version of the true cumulative distribution \( F \).

**Proposition 1.** Suppose \( X \) is an independent and identically distributed sequence with a continuous cumulative distribution function \( F \), then, in the limit as \( m \to \infty \)

\[
\hat{F}_{\omega,m}(c|X) \xrightarrow{a.s.} \phi_{\omega}(F(c)), \quad \text{and} \quad \tilde{F}_{\omega,m}(c) \xrightarrow{a.s.} \phi_{\omega}(F(c))
\]

(6)

where the probability weighting function \( \phi_{\omega} \) induced by the weight sequence \( \omega \) is given by

\[
\phi_{\omega}(x) = \sum_{i=1}^{n} \omega(i) \ U_{i:n}(x)
\]

(7)

and where \( U_{i:n} \) is the cumulative distribution function of the \( i^{th} \)-largest out of \( n \) independent draws from a uniform distribution on \([0,1]\),

\[
U_{i:n}(x) = \sum_{l=i}^{n} \binom{n}{l} x^l (1-x)^{n-l}
\]

(8)

**Proof.** See Appendix A. \( \square \)

Notice that while the proposition characterizes the limiting probability weighting functions only for the case of learning from i.i.d. data, the updating rule itself applies to dependent data just as well. In Section 3.3 below, we provide results on how the dependence structure of the data shapes the limiting probability distortions.

As \( U_{i:n} \) is the cumulative distribution function of the \( i \)-ranked draw among \( n \) draws from a uniform distribution, the so-called \( i^{th} \) order statistic, \( F_{i:n}(x) = U_{i:n}(F(x)) \) is the distribution of the \( i^{th} \) order statistic associated with drawing independently from \( F \).\(^7\) The intuition for the limiting, distorted empirical distribution is as follows: In the long run, all ranks from 1 to \( n \) appear with the same frequency. Whenever rank \( i \) appears, it receives weight \( \omega(i) \). Accordingly, in the long run, the empirical distribution converges to a mixture of the distributions of the order statistics where each rank is weighted as indicated by the rank-dependent attention weight function \( \omega \). In particular, \( \phi_{\omega} \) is a valid probability weighting function as defined, e.g., in Al-Nowaihi and Dhami (2010).

To gain a better insight on how the distortion acts upon the empirical distribution, we plot the limiting empirical distribution and the probability weighting function \( \phi_{\omega} \) in Figure 1. The

\(^7\)See David and Nagaraja (2004) for an introduction to the theory of order statistics which covers, e.g., the distribution of uniform order statistics.
parameters are set to the following values; \( n = 10, m = 1000, \omega(1) \propto 10, \omega(n) \propto 5 \) and all other values of omega are proportional to one. The objective distribution is standard normal. The left graph shows the objective cumulative distribution function, the subjective empirical distribution function as stated in Proposition 1 and the empirical distribution function obtained from a simulation using 1000 observations. The simulations indicate that the subjective empirical distribution function indeed converges to the derived functional form. The right graph shows the probability weighting function as given by Proposition 1 and compares it to objective probabilities. In the next Lemma we verify that the function \( \phi_{\omega} \) is indeed a valid probability weighting function.

**Lemma 1.** \( \phi_{\omega} \) is a probability weighting function, i.e., \( \phi_{\omega} : [0,1] \to [0,1] \) with \( \phi_{\omega}(0) = 0, \phi_{\omega}(1) = 1 \) and \( \phi_{\omega} \) is strictly increasing. \( \phi_{\omega} \) is a polynomial of degree at most \( n \) and thus a smooth function of \( x \). Moreover,

\[
\lim_{x \to 0} \frac{\phi_{\omega}(x)}{x} = n\omega(1), \quad \text{and} \quad \lim_{x \to 1} \frac{1 - \phi_{\omega}(x)}{1 - x} = n\omega(n). \tag{9}
\]

**Proof.** See Appendix A. \( \square \)

The two limits in (9) describe the weighting of very small or very large probabilities as discussed in Al-Nowaihi and Dhami (2010). The first limit shows that, in their terminology,
the probability weighting function $\phi_\omega$ finitely overweights infinitesimal probabilities if $\omega(1) > \frac{1}{n}$, i.e., if the smallest observations within an evaluation window get special attention. In contrast, if smallest observations tend to be ignored, $0 < \omega(1) < \frac{1}{n}$, $\phi_\omega$ positively underweights infinitesimal probabilities. Analogously, depending on whether $\omega(n)$ is smaller or larger than $\frac{1}{n}$, near-one probabilities are positively underweighted or finitely overweighted. We next study how the shape of the sequence $\omega$ influences the shape of the resulting probability weighting function. An increasing weight sequence means that higher ranked observations get larger weights. This results in a convex probability weighting function where all probabilities are overweighted. The converse holds for a decreasing weight sequence. Moreover, we provide sufficient conditions for $\phi_\omega$ to be $S$-shaped or inverse-$S$-shaped. Intuitively, a $U$-shape in the weight sequence $\omega$ means that high and low extreme observations are overweighted which should lead to an overwriting of extreme observations and thus to an inverse-$S$-shaped weighting function. In fact, we need to assume slightly more than a $U$-shape (i.e. unimodality) of $\omega$. We assume that the sequence is either convex or that it only takes three values. In the latter case, low-ranked, middle-ranked and high-ranked observations all get the same attention within each group and the middle-ranked get the least. We also provide an analogous result for inverse-$U$-shaped weighting functions.

**Lemma 2.** (i) If $\omega(1), \ldots, \omega(n)$ is increasing then $\phi_\omega$ convex.
(ii) If $\omega(1), \ldots, \omega(n)$ is decreasing then $\phi_\omega$ concave.
(iii) Suppose that either $\omega(1), \ldots, \omega(n)$ is convex or that $\omega(1), \ldots, \omega(n)$ is $U$-shaped and only takes three different values $\omega^l, \omega^m, \omega^h$. Then $\phi_\omega$ is inverse-$S$-shaped, i.e., there exists $x_0 \in [0, 1]$ s.t. $\phi_\omega$ is concave for $x \leq x_0$ and $\phi_\omega$ is convex for $x \geq 0$.
(iv) Suppose that either $\omega(1), \ldots, \omega(n)$ is concave or that $\omega(1), \ldots, \omega(n)$ is inverse-$U$-shaped and only takes three different values $\omega^l, \omega^m, \omega^h$. Then $\phi_\omega$ is $S$-shaped, i.e., there exists $x_0 \in [0, 1]$ s.t. $\phi_\omega$ is convex for $x \leq x_0$ and $\phi_\omega$ is concave for $x \geq 0$.

**Proof.** See Appendix A.

Notice that we do not rule out $x_0 = 0$ and $x_0 = 1$ in the lemma, thus allowing our $S$-shapes and inverse-$S$-shapes to be degenerate. Moreover, the conditions in (iii) and (iv) are, of course, only valid for $n \geq 3$ because it is not really meaningful to say that the sequence is convex or three-valued if it has less than 3 elements. Similarly, (i) and (ii) require $n \geq 2$ to be meaningful. If the weight sequence $\omega(1), \ldots, \omega(n)$ is constant, $\omega(i) = \frac{1}{n}$, it is both increasing and decreasing. In this case, the lemma implies that $\phi_\omega(x)$ is both convex and concave –
and thus linear. Accordingly, quite intuitively, we have $\phi_\omega(x) = x$ and thus no probability weighting. More typical cases of the probability weighting functions which are implied by different weight sequences are shown in Figure 2.

![Graphs showing functional forms of distortions related to four cases of $\omega(i)$ distinguished by Lemma 2.](image)

*Figure 2: Functional forms of distortions ($\phi_\omega$) related to the four cases of $\omega(i)$ distinguished by Lemma 2.*

### 3.2 A three-parameter model

Proposition 1 provides a microfoundation for a class of polynomial probability weighting functions with interpretable parameters. In this section, we explore this family further, focusing on a particularly tractable three-parameter version. To this end, we assume that only relative maxima and minima enter the learning process in a different way than all other observations within an evaluation window. Thus, we assume that $\omega(1)$ and $\omega(n)$ differ from $\omega(2), \ldots, \omega(n-1)$ while $\omega(i) = \omega(j)$ for $i, j = 2, \ldots, n-1$. Specifically, we consider the weight sequence

$$(\omega(1), \omega(2), \omega(3), \ldots, \omega(n)) = \left(\frac{1}{n} + \frac{n-2}{n} \theta \varepsilon, \frac{1}{n} + \varepsilon, \frac{1}{n}, \ldots, \frac{1}{n} + \frac{n-2}{n} (1 - \theta) \varepsilon\right)$$

(10)
where the parameter $\varepsilon \in [0, 1]$ controls the strength of overweighting of extremes\(^8\) while $\theta$ controls the relative weight of maxima vs. minima. In particular, for $\varepsilon = 1$ only maxima and minima are taken into account while for $\varepsilon = 0$ we have no distortion. In the next lemma, we collect some properties of the limiting probability weighting function associated with this specific sequence.

**Lemma 3.** For the weight sequence $\omega$ defined in (10) with $\theta, \varepsilon \in [0, 1]$ and $n \geq 3$, the associated limiting probability weighting function in the sense of Proposition 1 is given by

$$\phi_\omega(x) = (1 - \varepsilon)x + \varepsilon \theta x^n + \varepsilon (1 - \theta)(1 - (1 - x)^n).$$

(11)

For $\varepsilon > 0$ and $\theta \in (0, 1)$, this function $\phi_\omega$ is strictly inverse-S-shaped in the sense that it is strictly concave for $x < x_0$ and strictly convex for $x > x_0$ where the inflection point $x_0$ is given by

$$x_0 = \frac{\theta \frac{1}{n-2}}{\theta \frac{1}{n-2} + (1 - \theta) \frac{1}{n-2}}$$

(12)

*Proof.* See Appendix A. \(\square\)

The three-parameter functional form we have here is very flexible and can fit most common shapes of probability weighting functions. In Figure 3, we plot our function $\phi_\omega$ with $n = 10$, $\theta = 2/3$ and $\varepsilon = 1/2$ against a standard parametrization of Kahneman and Tversky (1979)’s famous probability weighting function\(^9\), showing that the two curves would be extremely hard to distinguish empirically. The parameter $\theta$ can be directly linked to the inflection point $x_0$ via the relation

$$\theta = \frac{(1 - x_0)^{n-2}}{x_0^{n-2} + (1 - x_0)^{n-2}}$$

that inverts (12). Empirically, inflection points tend to lie in the interval $[0, 1/2]$ which implies $\theta \in [1/2, 1]$ and thus $\omega(1) > \omega(n)$. In this case, maxima are thus overweighted to a greater extent than minima. In the limit $n \to \infty$, the function $\phi_\omega$ converges to a neo-additive probability weighting function (Chateauneuf et al., 2007) which is linear for $x \in (0, 1)$ while the case $n = 3$ corresponds to the cubic probability weighting functions studied, e.g., by Blavatskyy (2016). In his paper, Blavatskyy studies the relation between probability-weighted distorted distributions and the so-called $L$-moments of the undistorted distribution for the special case $n = 3$. The remaining results of this section are in a somewhat similar spirit. We

\(^8\)We could in principle also allow for negative $\varepsilon$ which corresponds to an underweighting of extremes. Technically, the analysis would be very similar but the interpretations would change.

\(^9\)Specifically, this function is given by $\phi(x) = \frac{x^\delta}{(x + (1-x)^\delta)^{1/\delta}}$ where $\delta = 0.6$. 

15
characterize how the subjective expected value of the distribution changes as the distortion as measured by $\varepsilon$ gets stronger. To this end, we derive a formula for the subjective expected value in terms of dispersion and skewness measures of the original distribution. These (new) dispersion and skewness measures are based on order statistics, similar to the $L$-moments studied by Blatavskyy.

**Proposition 2.** Consider a quantity of interest $X$ with cumulative distribution function $F$. Denote by $F_\varepsilon$ the distorted distribution function $F_\varepsilon(x) = \phi_\omega(F(x))$ where $\phi_\omega$ is given in (11). Then we have the relation

$$E_\varepsilon[X] = E[X] + \varepsilon s_n(X) \left( \theta - \frac{1}{2} \right) + \varepsilon s_n(X) \gamma_n(X)$$

(13)

where $E_\varepsilon$ and $E$ denote the subjective expected value under $F_\varepsilon$ and the objective expected value under $F$. The quantities $s_n(X)$ and $\gamma_n(X)$ are defined as

$$s_n(X) = E[X_{1:n}] - E[X_{n:n}] \quad \text{and} \quad \gamma_n(X) = \frac{\frac{1}{2}E[X_{1:n}] + \frac{1}{2}E[X_{n:n}] - E[X]}{E[X_{1:n}] - E[X_{n:n}]}$$

where $X_{1:n}$ and $X_{n:n}$ are distributed like the maximum and the minimum of $n$ i.i.d. draws from $F$. Moreover, $s_n(X)$ and $\gamma_n(X)$ are, respectively, valid measures of scale and of skewness in the sense of Oja (1981) and Groeneveld and Meeden (1984).
Proof. See Appendix A.

The fact that $s_n$ and $\gamma_n$ are valid measures of scale and skewness means that, essentially, they can be interpreted in the same way as the usual moment-based standard deviation and skewness.\(^{10}\) Equation (13) captures how different properties of the weighting function and the distribution of $X$ lead to an upward or downward distortion in the subjective expected value of $X$. The middle term involving $\theta$ simply captures that, ceteris paribus, if more weight is put on the largest observations, $\theta > 1/2$, this leads to an upward bias in the perceived mean. Similarly, we obtain a downward bias for $\theta < 1/2$. These effects are stronger if $X$ is more volatile so that large and small observations differ more strongly (large $s_n(X)$) or if the distortion is stronger (large $\varepsilon$). The last term implies an additional upward bias whenever $X$ is positively skewed (right skewed) and, conversely, a downward bias under negative skewness. Intuitively, overweighting of extremes leads to a distortion towards the more extreme tail. Under positive skewness, we thus see an upward distortion in the subjective mean, and similarly a downward distortion for negative skewness. This is in line with the literature on skewness seeking behavior (Barberis & Huang, 2008; Ebert & Wiesen, 2011; Barberis et al., 2016).

We next discuss how overweighting of extremes distorts the subjective scale of the distribution as measured by variance. For ease of exposition, we assume that $X$ is centered, $E[X] = 0$, and that the distribution of $X$ is symmetric around its mean. Note that for any quantity $h(X)$ that depends on $X$ the distorted expectation can be related to expectations of $h$ applied to order statistics analogously with (13), see also the proof of Proposition 2,

$$E_{\varepsilon}[h(X)] = (1 - \varepsilon)E[h(X)] + \varepsilon \theta E[h(X_{1:n})] + \varepsilon(1 - \theta)E[h(X_{n:n})].$$

Using that symmetry implies $-E[X_{1:n}] = E[X_{n:n}]$ and $E[X_{1:n}^2] = E[X_{n:n}^2]$, it immediately follows that

$$E_{\varepsilon}[X] = \varepsilon(2\theta - 1)E[X_{1:n}], \quad \text{and} \quad E_{\varepsilon}[X^2] = (1 - \varepsilon)E[X^2] + \varepsilon E[X_{1:n}^2]$$

and thus

$$\text{Var}_{\varepsilon}(X) = \text{Var}(X) + \varepsilon(E[X_{1:n}^2] - \text{Var}(X)) - \varepsilon^2(2\theta - 1)^2E[X_{1:n}^2].$$

\(^{10}\)Intuitively, it is not necessary to take the square of a random variable to decide how spread out it is. There can thus be many different measures of scale like the mean absolute deviation $E[|X - E[X]|]$ or the quantity $s_n$ that appears naturally in our setting. Similarly, the question whether $X$ is left or right-skewed is not intrinsically a question about $X^3$. There is thus room for many different skewness measures. See the proof of Proposition 2 or Groeneveld and Meeden (1984) for the technical details.
For small distortions, \( \varepsilon \approx 0 \), we can ignore the second term which is of order \( \varepsilon^2 \). We see that \( \text{Var}_\varepsilon(X) \) is increasing in \( X \) for small \( \varepsilon \) as \( E[X_{1:n}^2] - \text{Var}(X) = E[X_{1:n}^2] - E[X^2] > 0 \). If the attention distortions are symmetric, \( \theta = \frac{1}{2} \), the lower order is always zero and \( \text{Var}_\varepsilon(X) \) is increasing in \( \varepsilon \). Due to symmetry, distortions move probability mass equally into both tails and thus increase perceived variability. However, \( \text{Var}_\varepsilon(X) \) is decreasing in \( |\theta - \frac{1}{2}| \) which is a measure of the asymmetry of the belief distortion. In the extreme case \( \varepsilon = 1 \) where attention is fully concentrated on extremes, the subjective variance becomes

\[
\text{Var}_1(X) = E[X_{1:n}^2] - (2\theta - 1)^2 E[X_{1:n}]^2.
\]

which decreases from \( E[X_{1:n}^2] > \text{Var}(X) \) to \( \text{Var}(X_{1:n}) \) as \( \theta \) moves away from \( \frac{1}{2} \) to one of the extremes, \( \theta \in \{0, 1\} \). For many distributions of interest, the variance of the maximum \( \text{Var}(X_{1:n}) \) is smaller than the objective variance \( \text{Var}(X) \). Thus, sufficiently asymmetric belief distortions can lead to a decrease in subjective risk perceptions. Intuitively, attention is diverted into one of the tails. Moreover, conditionally on coming from this tail, observations are less variable than unconditionally. For example, if \( X \) has full support on a bounded interval \([-c, c]\) then, for sufficiently large \( n \), \( X_{1:n} \) will be close to \( c \) with high probability and \( \text{Var}(X_{1:n}) \) converges to 0 as \( n \) grows.

### 3.3 Dependent Realizations

Our theory of inattentive learning is intuitively applied in a setting of an investor learning about asset price returns. Though, the learning dynamics can be applied more generally. Our results above are derived under the assumption that the underlying process generates i.i.d. observations\footnote{Note that this assumption concerns the objective data generating process and does not impose any restrictions for the subjective beliefs. We do however implicitly assume that if the agent had unlimited attention he would give all observations equal attention weights. Such learning dynamics is rational if the investor assumes the realizations are i.i.d. or in case the agent is interested in learning the unconditional distribution. (Landier et al., 2019) show that individuals find it very difficult to distinguish i.i.d. observations from correlated observations, we therefore leave more detailed assumptions by the agent to further work.}, however, this assumption is not always applicable. To show the results more generally we investigate in this section the effects of inattentive learning when the underlying distribution exhibits a non-zero dependence structure. We address the concern one may have that inattentive learning may generate very different belief dynamics if the underlying data generating process are not i.i.d. We therefore adapt the process \( X \) to a process \( Y \), by changing...
the data generating process to an AR(1) process with persistence $\rho$.

$$ Y_t = \rho Y_{t-1} + \sqrt{1 - \rho^2} X_t $$

where $X$ is standard normal and i.i.d. distributed over time. $Y_t$ is also standard normally distributed with autocorrelation $\rho$, such that for $\rho$ equals zero both process are equal. Under this new data generating process the results of the previous sections are difficult to prove. Instead, we perform simulations and compare the belief dynamics for values of $\rho$ between zero and one.

![Figure 4: The figures show the subjective mean, volatility, sharpe ratios and skewness over time for the simple inattentive learning dynamics applied to a standard normal process with autocorrelation $\rho$. The upper graphs show the median over 100,000 simulations for intermediate points in time. The lower graphs show the distance between of the 99-percentile and the 1-percentile. The attention parameters are set to $n = 10$, $\varepsilon = \frac{1}{2}$ and $\theta = \frac{2}{3}$.](image)

For each value of $\rho$ between -1 and 1, we perform 100,000 simulations each with a time length of 2000 observations. The graphs in Figure 4 show at intermediate points in time the subjective beliefs plotted over $\rho$ of an agent that exhibits the baseline simple limited attention.

The top left graph shows that for autocorrelations fairly close to zero do not generate different subjective means. In case the autocorrelation comes close to minus one and one,
the bias in the subjective mean plunges. The corresponding accuracies as measured by the distance between the 99th percentile and the 1 percentile are shown in the lower left graph. It shows that the precision of the subjective mean for any point in time is lower for larger autocorrelation and that learning is faster (higher precision) for smaller autocorrelations. The subjective volatility displays a subjective bias that is of similar shape as the bias in the subjective mean, though, note that the objective volatility equals one. The graphs of the precision of subjective volatility are fairly intuitive. For autocorrelations near one or minus one, the agent will observe either always the same values or always very different values. In the first case the agent is likely to see relatively few extremes and therefore has a relatively high precision. However, when the process always takes very different values, the agent observes many extreme realizations and has therefore low precision. As the number of observations tends to infinity, the precision of volatility becomes symmetric with respect to the autocorrelation, $\rho$. The bias in volatility is smallest for zero autocorrelation and increases when the autocorrelation diverges from zero. For non-zero autocorrelations the asymmetric nature of the inattentive learning generates lower precision. Combining the subjective mean and the subjective volatility we obtain that the subjective sharpe ratios are relatively stable for different autocorrelations and over the number of observations. Again, for autocorrelations close to zero the bias in the sharpe ratio is very similar. Interestingly the subjective sharpe-ratio is also very stable over the number of observations. The precision of the subjective sharpe ratio is mostly driven by the precision of the subjective mean and therefore decreases with the autocorrelation for most of the parameter spectrum. The bias of skewness decreases for all autocorrelation values and acts relatively similar for autocorrelations close to zero. For autocorrelations near one the bias of the subjective skewness drops to zero. Just as for the other moments, the reverse relation between thee bias and the precision is also present for skewness.

Figure 4 shows that the results for the mean, volatility and skewness of the subjective distribution are fairly similar for a relatively large range of autocorrelations around zero. In case the autocorrelation equals values near one or minus one, the results tend to change such that the beliefs become less biased, though the precision of these beliefs is larger and therefore the agent is learning more slowly. In the setting of daily returns, the autocorrelations are strongly significant, though the empirically found range leads to results relatively similar to the i.i.d. assumption French and Roll (1986). Our findings above therefore indicate that our theoretical results will also approximately hold when applied to daily returns. Additionally, our results also indicate that for setting where the autocorrelation is relatively large in absolute
magnitude, the results may be different.

4 An Empirical Test  

In the previous we proposed a theory of inattentive learning and we derived it’s implications for the subjective belief distribution of the agent. Our theory is based upon the results of survey expectations, though, for establishing it’s practical relevance we use financial market data and test for significant impact of inattentive learning. In our analysis we use the CRSP database for data on individual US stocks from 1926 up to 2018. We follow Pástor and Stambaugh (2003) and trim the sample by imposing three requirements. We only include stocks that are listed at the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX), are classified as ordinary common shares (CRSP share codes 10 and 11) and of which the stock price in the month preceding the observation was between 5 and 1000 dollars. In our analysis make use of the Fama-French (Fama & French, 1993) factors, which we download from Kenneth French’s website.

4.1 Cross-sectional Variation

We are interested in the extent to which agents in financial markets are subject to inattentive learning. Our results using survey data suggest that at least a significant proportion of the agents are subject to inattentive learning. Similar to Barberis et al. (2016), we argue that such agents will drive up the price of stocks that have a particular high subjective value according to inattentive learning. The intuition is as follows. If agents are subject to inattentive learning, their subjective beliefs will differ from objective beliefs. In the cross-section of stocks there will be stocks of which the subjective beliefs are more favorable than for other stocks. The agents will therefore buy relatively more of the stocks with favorable subjective beliefs and relatively less of stocks with less favorable subjective beliefs. When a significant part of the participants on financial markets act along these lines, the stocks with favorable subjective beliefs will become overvalued and the less favorable stocks become undervalued. Subsequently, the overvalued stocks earn relatively low returns and the undervalued stocks earn relatively high returns. We will use the Fama and MacBeth (1973) framework to identify this effect.

Barberis et al. (2016) study the effect of static prospect theory estimated on a rolling basis, where they perform a cross-sectional analysis and compare subsequent returns of stocks with low prospect theory value versus stocks with high prospect theory value.
From the theoretical results we know how the subjective mean is influenced by objective moments. Therefore, we could in principle perform our analysis on subsamples of stocks based on differences in scale and skewness. However, we argue that finding significant effects including all stocks shows a greater importance of inattentive learning. The analysis we will therefore not be restricted to some subset of the firms in the CRSP database.

In the verification section we found strong significance of daily returns on survey expectations, we therefore also use daily returns for the analysis in this section. The inattentive learning dynamics will follow the simplified three-parameter setting with the parameters $n = 10$, $\varepsilon = \frac{1}{2}$ and $\theta = \frac{2}{3}$. Additionally we set the memory window to 30 days. Then, for each stock we compute the subjective belief distribution over time on a daily frequency. As we cannot compare entire belief distributions among assets, we choose to compare it’s mean, variance and skewness. Following our intuition above, we expect that stocks that are relatively more favorable will be overvalued compared to less favorable stocks. As our theory does not assume any specific preferences, we make a relatively general assumption that the subjective sharpe ratio proxies whether investors find a stock more favorable than another. We sort stocks in decile portfolios based on their subjective sharpe ratio and derive the next periods excess returns and alpha’s when controlling for commonly used risk factors.

Table 3 shows the results of the daily analysis of inattentive learning when using the sharpe ratio as a measure of favorability. The excess returns are monotonically decreasing when shifting from low favorability to high favorability. This is in line with the intuition above that if agents are subject to inattentive learning the stocks they find more favorable will become overvalued and therefore earn low subsequent returns. This pattern persists when also including the momentum factor (Carhart, 1997) and the liquidity factor (Pástor & Stambaugh, 2003). In the last specification of Table 3 we determine whether the effect we find is similar to the effect caused by prospect theory. We therefore compute the measure of Barberis et al. (2016) on a daily level using the same memory window as used for our measure of inattentive learning. Also for this specification there is a strong relation between the returns and the favorability induced by limited attention. We note that the annualized daily returns in Table 3 are very extreme. This is due to the high frequency of sorting used to generate these results. An investing strategy replicating these results would incur very high trading costs due to the daily sorts and would therefore have returns that will be a lot lower. Nevertheless, our results in Table 3 do highlight that inattentive learning generates misvaluation in the stock market.

To give a better impression on the magnitude of the effects of Table 3, we forecast monthly
returns instead of daily returns. The monthly frequency is more commonly used in studies using the Fama and MacBeth (1973) framework, which lowers the degree of trading costs. The portfolio sorting is identical to our results above, however, now we perform our analysis using the subsequent monthly returns for each sorted portfolio. The results are in Table 4. The monthly regressions show low though significant returns relative to the daily results. As we are using fairly little information to forecast a relatively long maturity return it is logical that the effects will be less strong. The increase in maturity also causes a fading of the monotonicity in our results over the favorability of stocks. However, we do see significant difference between the outer percentiles in terms of returns for almost all specifications, which indicates that

Table 3: The results show annualized excess returns and alphas of daily returns of decile portfolios sorted from least favorable (P1) by means of limited attention up to most favorable (P10). The four-factor model adds the momentum factor by Carhart (1997) and the five-factor model adds the liquidity factor by Pástor and Stambaugh (2003). BMW represents the factor of prospect theory by Barberis et al. (2016), where we sort daily similar to our measure of limited attention and set the evaluation window also to the same length. The proxy is constructed using the difference between the top and bottom sort-decile. Results are very similar when choosing the top 30 percent minus the bottom 30 percent instead. Newey and West (1987)-adjusted t-statistics are between brackets.
Table 4: The results show annualized excess returns and alphas of monthly returns of decile portfolios sorted from least favorable (P1) by means of limited attention up to most favorable (P10). The four-factor model adds the momentum factor by Carhart (1997) and the five-factor model adds the liquidity factor by Pástor and Stambaugh (2003). BMW represents the factor of prospect theory by Barberis et al. (2016), where we sort daily similar to our measure of limited attention and set the evaluation window also to the same length. The proxy is constructed using the difference between the top and bottom sort-decile. Results are very similar when choosing the top 30 percent minus the bottom 30 percent instead. Newey and West (1987)-adjusted t-statistics are between brackets.

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P1-P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Return</td>
<td>EW</td>
<td>0.1630</td>
<td>0.1406</td>
<td>0.1549</td>
<td>0.1542</td>
<td>0.1574</td>
<td>0.1208</td>
<td>0.1187</td>
<td>0.1001</td>
<td>0.0824</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.31)</td>
<td>(4.51)</td>
<td>(5.27)</td>
<td>(5.12)</td>
<td>(4.61)</td>
<td>(4.28)</td>
<td>(4.08)</td>
<td>(3.48)</td>
<td>(3.04)</td>
</tr>
<tr>
<td></td>
<td>VW</td>
<td>0.1651</td>
<td>0.1347</td>
<td>0.1567</td>
<td>0.1817</td>
<td>0.1914</td>
<td>0.1249</td>
<td>0.0533</td>
<td>0.1106</td>
<td>0.1013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.38)</td>
<td>(4.06)</td>
<td>(3.88)</td>
<td>(5.10)</td>
<td>(4.20)</td>
<td>(3.77)</td>
<td>(1.58)</td>
<td>(3.56)</td>
<td>(2.68)</td>
</tr>
<tr>
<td>3-Factor EW</td>
<td>alpha</td>
<td>0.1160</td>
<td>0.0917</td>
<td>0.1095</td>
<td>0.1053</td>
<td>0.1098</td>
<td>0.0698</td>
<td>0.0672</td>
<td>0.0500</td>
<td>0.0323</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.17)</td>
<td>(3.41)</td>
<td>(4.16)</td>
<td>(3.96)</td>
<td>(3.47)</td>
<td>(2.80)</td>
<td>(2.64)</td>
<td>(1.97)</td>
<td>(1.33)</td>
</tr>
<tr>
<td></td>
<td>VW</td>
<td>0.1166</td>
<td>0.0784</td>
<td>0.0968</td>
<td>0.1328</td>
<td>0.1492</td>
<td>0.0822</td>
<td>0.0189</td>
<td>0.0461</td>
<td>0.0439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.99)</td>
<td>(2.54)</td>
<td>(2.69)</td>
<td>(3.93)</td>
<td>(3.28)</td>
<td>(2.64)</td>
<td>(0.56)</td>
<td>(1.46)</td>
<td>(1.30)</td>
</tr>
<tr>
<td>4-Factor EW</td>
<td>alpha</td>
<td>0.0291</td>
<td>-0.0136</td>
<td>0.0181</td>
<td>0.0210</td>
<td>0.0120</td>
<td>-0.0107</td>
<td>-0.0136</td>
<td>-0.0266</td>
<td>-0.0421</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.90)</td>
<td>(-0.38)</td>
<td>(0.51)</td>
<td>(0.61)</td>
<td>(0.34)</td>
<td>(-0.31)</td>
<td>(-0.39)</td>
<td>(-0.75)</td>
<td>(-1.18)</td>
</tr>
<tr>
<td></td>
<td>VW</td>
<td>0.1934</td>
<td>0.2824</td>
<td>0.1201</td>
<td>0.1777</td>
<td>0.2705</td>
<td>0.1738</td>
<td>-0.0793</td>
<td>-0.0107</td>
<td>0.1164</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.48)</td>
<td>(3.73)</td>
<td>(0.70)</td>
<td>(2.81)</td>
<td>(2.48)</td>
<td>(1.80)</td>
<td>(-0.76)</td>
<td>(-0.09)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>5-Factor EW</td>
<td>alpha</td>
<td>0.0367</td>
<td>-0.0103</td>
<td>0.0255</td>
<td>0.0269</td>
<td>0.0173</td>
<td>-0.0064</td>
<td>-0.0111</td>
<td>-0.0242</td>
<td>-0.0378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.19)</td>
<td>(-0.31)</td>
<td>(0.77)</td>
<td>(0.82)</td>
<td>(0.51)</td>
<td>(-0.20)</td>
<td>(-0.34)</td>
<td>(-0.71)</td>
<td>(-1.12)</td>
</tr>
<tr>
<td></td>
<td>VW</td>
<td>0.2175</td>
<td>0.2517</td>
<td>0.1885</td>
<td>0.2441</td>
<td>0.2866</td>
<td>0.2753</td>
<td>0.0162</td>
<td>0.1275</td>
<td>0.1530</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.80)</td>
<td>(3.53)</td>
<td>(1.24)</td>
<td>(3.84)</td>
<td>(2.64)</td>
<td>(3.24)</td>
<td>(0.14)</td>
<td>(1.10)</td>
<td>(1.34)</td>
</tr>
<tr>
<td>BMW EW</td>
<td>-0.0549</td>
<td>-0.0675</td>
<td>-0.0453</td>
<td>-0.0532</td>
<td>-0.0647</td>
<td>-0.0671</td>
<td>-0.0726</td>
<td>-0.0840</td>
<td>-0.0973</td>
<td>-0.0975</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.12)</td>
<td>(-2.06)</td>
<td>(-1.45)</td>
<td>(-1.78)</td>
<td>(-2.04)</td>
<td>(-2.20)</td>
<td>(-2.34)</td>
<td>(-2.60)</td>
<td>(-3.03)</td>
</tr>
<tr>
<td></td>
<td>VW</td>
<td>-0.0344</td>
<td>-0.0318</td>
<td>-0.0248</td>
<td>-0.0085</td>
<td>-0.0297</td>
<td>-0.0499</td>
<td>-0.0830</td>
<td>-0.0256</td>
<td>-0.0928</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-1.13)</td>
<td>(-0.82)</td>
<td>(-0.58)</td>
<td>(-0.24)</td>
<td>(-0.76)</td>
<td>(-1.19)</td>
<td>(-1.93)</td>
<td>(-0.73)</td>
<td>(-2.44)</td>
</tr>
</tbody>
</table>

The effect of inattentive learning is still significantly present at the one month horizon. The magnitude is also still economically sizable such that the replicating portfolio would generate also a positive alpha after trading costs.

Up to this point we only used the subjective Sharpe ratio as a measure of favorability of a stock. Next, we show results for other characteristics of the subjective distribution, like the mean, volatility and skewness. In Table 5 we shows the results of going long the low decile and short the top decile, and we perform this analysis on both the daily returns as the monthly returns.

First note that the sign of the volatility results are mainly negative, as we look at a
Table 5: The results are from decile rankings on either the subjective mean, volatility or skewness and going long the least favorable stocks and long the most favorable stocks. BMW-10 shows results when controlling for a decile long-short portfolio based on the prospect theory measure of (Barberis et al., 2016), and BMW-3 when going long the top 30 percent and short the bottom 30 percent.

portfolio of long low subjective volatility stocks and short high volatility stocks. Then we observe that the results for daily return predictions are highly significant, though the monthly return results show barely significant effects. This is due to the high frequency of changes in subjective beliefs which therefore does not seem to give a lot of explanatory power on the monthly horizon, apart from our earlier findings using the sharpe ratio. Our results therefore provide evidence that investors indeed act along the lines of inattentive learning, albeit on a relatively short-horizon which induces short-term predictability to a greater extent than longer-term predictability.

5 Conclusion

In this paper we have several contributions to the literature. Our main contribution is that we propose a model of inattentive learning of which we can derive the limiting belief distortion, the subjective mean and subjective variance in closed form. Our solutions show that inattentive
learning is able to generate familiar belief distortions like, probability weighting, a preference for positive skewness, overestimation of volatility and overextrapolation of returns. These belief distortions are widely used in financial economics, though, they are commonly assumed to have different and exogenous origins. We show that a model of inattentive learning is able to generate these belief distortions simultaneously.

Another contribution of our model is that it generates a new class of probability weighting functions with interpretable parameters. The model provides a direct link between the parameterization of the inattentive learning to the limiting probability weighting distortion. We also provide intuition on how different parameterizations of the model lead to different belief distortions.

We also perform a verification of inattention. Following Greenwood and Shleifer (2014), who show that survey data is a good measure of beliefs on financial markets, we show that these beliefs are formed by investors who allocate more attention towards extreme realizations. Additionally we show that when forming beliefs on a frequent basis, investors tend to memorize 30 up to at most 50 lagged daily returns. (Greenwood & Shleifer, 2014) show that these survey expectations strongly correlate with financial market beliefs, which implies that a significant proportion of the participants in financial market do exhibit inattentive learning. This emphasizes the relevance of our model.

In the last chapter we provide empirical evidence of inattentive learning using stock market data. We find evidence for inattentive learning via several characteristics of the subjective belief distribution for individual assets. Inattentive learning drives up the price of stocks that perceived as more favorable due to high attention such that these stocks perform poorly shortly afterwards.
References


Appendix A - Proofs

Proof of Proposition 1. We start by showing the result for $\tilde{F}_{w,m}$. To this end, we consider $m \gg n$ decompose the sum in the definition of $\tilde{F}_{w,m}$ into $n + 1$ separate sums, using that due to the finite window length $n$, observation weights which lie more than $n$ time steps apart are independent. Specifically, defining $\gamma_k(c) \equiv w(\rho_n(k, X)1_{\{X_k \leq c\}})$, we write

$$\tilde{F}_{w,m}(c, x) = \frac{n}{m} \sum_{k=1}^{m} w(\rho_n(k, X))1_{\{X_k \leq c\}} = \frac{n}{m} \sum_{k=1}^{m} \gamma_k(c) = S_0 + S_1^{(m)} + S_2^{(m)} + \ldots + S_n^{(m)}$$

(15)

where

$$S_0 = \frac{n}{m} \sum_{i=1}^{n} \gamma_i(c) \quad \text{and} \quad S_j^{(m)} = \frac{n}{m} \sum_{k \in I_j(m)} \gamma_k(c)$$

and where the index sets are given by $I_j(m) \equiv \{i \geq n, i \mod n = j\}$. Thus, the first $n$ steps (where the updating is defined differently) are separated from the rest in the sum $S_0$. Clearly, $S_0$ converges to 0 as $m$ goes to infinity.\(^{13}\) The remaining sums $S_j^{(m)}$ consist of terms at a distance $n$, i.e.,

$$S_j^{(m)} = \frac{n}{m} (\gamma_{n+j}(c) + \gamma_{2n+j}(c) + \gamma_{3n+j}(c) \ldots)$$

Note that each sum $S_j^{(m)}$ is a sum of i.i.d. terms which becomes infinite as $m$ goes to infinity. Moreover, $S_j^{(m)}$ and $S_k^{(m)}$ are identically distributed (but, of course, not independent). To conclude the argument for $\tilde{F}_{w,m}$, it thus suffices to show that

$$S_j^{(m)} \xrightarrow{a.s.} \frac{1}{n} \sum_{i=1}^{n} \omega(i)U_{i:n} \left( F(c) \right) = \frac{1}{n} \phi_w \left( F(c) \right)$$

for all $j = 1, \ldots, n$. We can write $S_j^{(m)}$ as

$$S_j(m) = \frac{n}{m} \#I_j(m) \left( \frac{1}{\#I_j(m)} \sum_{k \in I_j(m)} \gamma_k(c) \right)$$

(16)

where the leading, deterministic factor $\frac{n}{m} \#I_j(m)$ clearly converges to 1 as $m$ grows. The final term in brackets is now an i.i.d sum of $\#I_j(m)$ terms which are bounded by $\max_i w(i)$ where $\#I_j(m)$ is deterministic and goes to infinity as $m$ goes to infinity. Thus, the law of large numbers applies,\(^{14}\) showing that the term in brackets converges to $E[\gamma_k(c)]$ where $k \geq n + 1$

\(^{13}\)This implies in particular that it is actually irrelevant for the limit how we initialize the updating in the first $n$ steps.

\(^{14}\)We can, for instance, apply the version in Kallenberg (2006), Theorem 4.23.
is arbitrary. To compute $\mathbb{E}[\gamma_k(c)]$, we write

$$
\mathbb{E}[\gamma_k(c)] = \mathbb{E}\left[w(\rho_n(k,X))1_{\{X_k \leq c\}}\right]
$$

$$
= \sum_{i=1}^{n} \mathbb{P}(\rho_n(k,X) = i) \mathbb{E}\left[w(\rho_n(k,X))1_{\{X_k \leq c\}} | \rho_n(k,X) = i\right]
$$

$$
= \sum_{i=1}^{n} \frac{1}{n}w(i) \mathbb{E}\left[1_{\{X_k \leq c\}} | \rho_n(k,X) = i\right]
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} w(i) \mathbb{P}(X_k \leq c | \rho_n(k,X) = i)
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} w(i) F_{i,n}(c)
$$

Here, the first step used the law of iterated expectations. The second step used that the rank of $X_k$ within its observation window is uniformly distributed a priori and that $w(\rho_n(k,X))$ is known conditional on this rank. In the last two steps, we simply observe that the probability in the formula corresponds to the distribution function of the $i$th order statistic. Inserting $F_{i,n}(c) = U_{i,n}(F(c))$ completes the argument. The convergence of $\tilde{F}_{w,m}$ is thus settled and it remains to extend the argument to $\hat{F}_{w,m}$. To this end, note that the previous argument with $c = \infty$ (i.e. without the factors $1_{\{X_k \leq c\}}$) implies that

$$
\frac{n}{m} \sum_{k=1}^{m} w(\rho_n(k,X)) \overset{a.s.}{\longrightarrow} 1
$$

and thus also

$$
\frac{1}{n} \sum_{k=1}^{m} w(\rho_n(k,X)) \overset{a.s.}{\longrightarrow} 1
$$

Therefore,

$$
\hat{F}_{w,m}(c|X) = \frac{\hat{F}_{w,m}(c|X)}{\frac{n}{m} \sum_{k=1}^{m} w(\rho_n(k,X))} \overset{a.s.}{\rightarrow} \phi_w(F(c))
$$

Proof of Lemma 1. First, recall the representation of the weighting function in terms of the cumulative distribution functions of uniform order statistics $U_{i,n}$,

$$
\phi_w(x) = \sum_{i=1}^{n} w(i) U_{i,n}(x).
$$
Thus, the function $\phi_w$ is a convex combination of the functions $U_{i,n}$ which are strictly increasing with $U_{i,n}(0) = 0$ and $U_{i,n}(1) = 1$ as is shown, e.g., in David and Nagaraja (2004). Thus, these properties are inherited by $\phi_w$. Similarly, $\phi_w$ inherits the property of being a polynomial of degree at most $n$ from the $U_{i,n}$. To compute the limits for $x \to 0$ and $x \to 1$, note that both the denominators and the numerators vanish so we rely on L’Hôpital’s rule. To this end, note that the first derivative of the weighting function can be expressed in terms of the density functions $u_{i,n}^n$ of the uniform order statistics,

$$
\phi'_w(x) = \sum_{i=1}^n w(i) u_{i,n}(x) = \sum_{i=1}^n w(i) n\binom{n-1}{i-1} x^{i-1}(1-x)^{n-i}
$$

$$
= \sum_{k=0}^{n-1} w(k+1) n\binom{n-1}{k} x^k(1-x)^{(n-1)-k}
$$

(20)

where the formula for $u_{i,n}$ is found, e.g., in David and Nagaraja (2004). Thus, we conclude that

$$
\lim_{x \to 0} \frac{\phi_w(x)}{x} = \phi'_w(0) = nw(1)
$$

and

$$
\lim_{x \to 1} \frac{1 - \phi_w(x)}{1-x} = \phi'_w(1) = nw(n).
$$

Proof of Lemma 2. To study the shape $\phi_w$, we first compute its second derivative, starting with the first derivative stated in (20).

$$
\phi''_w(x) = \sum_{k=1}^{n-1} w(k+1) n\binom{n-1}{k} k x^{k-1}(1-x)^{(n-1)-k}
$$

$$
- \sum_{k=0}^{n-2} w(k+1) n\binom{n-1}{k} (n-1-k)x^k(1-x)^{(n-2)-k}
$$

$$
= \sum_{k=0}^{n-2} w(k+2) n\binom{n-1}{k+1} (k+1)x^k(1-x)^{(n-2)-k}
$$

$$
- w(k+1) n\binom{n-1}{k} (n-1-k)x^k(1-x)^{(n-2)-k}
$$

$$
= \sum_{k=0}^{n-2} (w(k+2) - w(k+1)) n(n-1)\binom{n-2}{k} x^k(1-x)^{(n-2)-k}
$$

where the final step used that the binomial coefficients satisfy

$$
(k+1)\binom{n-1}{k+1} = (n-1-k)\binom{n-1}{k} = (n-1)\binom{n-2}{k} = \frac{(n-1)!}{k!(n-2-k)!}.
$$
Thus, we can conclude that if the sequence $w$ is increasing, $w(k+2) \geq w(k+1)$, then the weighting function $\phi_w$ is convex as its second derivative is positive. This proves (i) and, by analogy, (ii). To show (iii) and (iv), we first find by an analogous calculation that\(^{15}\)

$$
\phi''_w(x) = \sum_{k=0}^{n-3} (w(k+3) - 2w(k+2) + w(k+1)) \frac{n(n-1)(n-2)}{k} x^k (1-x)^{(n-3)-k}.
$$

Thus, if $w$ is convex, $w(k+3) - 2w(k+2) + w(k+1) \geq 0$ then $\phi''_w(x)$ is positive and $\phi''_w(x)$ is increasing in $x$. Thus, $\phi''_w(x)$ changes signs at most once, and if it does then from negative to positive. Thus, $\phi_w$ can change at most once from being locally concave to being locally convex. This proves the claim for convex $w$ in (iii). The claim for concave $w$ in (iv) follows analogously. To prove (iii) and (iv), it remains to show that $\phi''_w$ switches signs at most once when $w$ takes only three values as described. In this case, there exist $k$ and $j$, $j > k$, such that $w(k+2) - w(k+1) = w^m - w^l$ and $w(j+2) - w(j+1) = w^h - w^m$ while all other increments in the weight sequence are zero. Thus, the second derivative can be written as

$$
\phi''_w(x) = (w^m - w^l)n(n-1)\binom{n-2}{k} x^k (1-x)^{(n-2)-k}
+ (w^h - w^m)n(n-1)\binom{n-2}{j} x^j (1-x)^{(n-2)-j}
= x^k (1-x)^{n-j}(A_k(1-x)^{j-k} + A_jx^{j-k})
$$

where $A_k = (w^m - w^l)n(n-1)\binom{n-2}{k}$ and $A_j = (w^h - w^m)n(n-1)\binom{n-2}{j}$. In case (iii), we know that $A_k < 0$ and $A_j > 0$. Thus, both $A_k(1-x)^{j-k}$ and $A_jx^{j-k}$ are increasing in $x \in [0,1]$, implying that their sum is increasing and thus switches from being negative to being positive at most once. Thus $\phi_w$ can change at most once from being locally concave to being locally convex. The argument in case (iv) is analogous.\hfill \square

**Proof of Lemma 3.** We begin by establishing the connection between the weight sequence and the probability weighting function. To this end, recall first that the cumulative distribution functions $U_{i,n}$ of uniform order statistics satisfy $U_{1,n}(x) = x^n$, $U_{n,n}(x) = 1 - (1 - x)^n$ and $U(x) = \frac{1}{n} \sum_{i=1}^{n} U_{i,n}(x)$ where $U(x) = x$ is the uniform cumulative distribution function\(^{16}\) and

\(^{15}\) Indeed, the calculation that derives $\phi''_w$ from $\phi'_w$ is exactly the same as the one that lead from $\phi'_w$ to $\phi''_w$ if one redefines $\hat{w}(k) = w(k+1) - w(k)$ and $\hat{n} = n - 1$.

\(^{16}\) The last, well-known identity captures that drawing from $U$ is the same as first drawing a rank $I$ uniformly from 1 to $n$ and then drawing from $U_{I,n}$.  

32
that our limiting probability weighting functions are given by

\[ \phi_w(x) = \sum_{i=1}^{n} w(i) U_{i:n}(x). \]  

(21)

Consider the following weight sequence which gives extra weight \( A \) and \( B \) to the extreme observations

\[ (w(1), w(2), w(3), \ldots, w(n)) = \left( \frac{1}{n} + A, \frac{1}{n} - \frac{A + B}{n - 2}, \frac{1}{n} - \frac{A + B}{n - 2}, \ldots, \frac{1}{n} + B \right). \]  

(22)

Then, starting from (21) and using the connection between \( U(x) \) and the \( U_i:n \), we can write the limiting weighting function as

\[ \phi_w(x) = U(x) + A U_{1:n}(x) + B U_{n:n}(x) - \frac{A + B}{n - 2} \sum_{i=2}^{n-2} U_{i:n}(x) \]

\[ = U(x) + A U_{1:n}(x) + B U_{n:n}(x) - \frac{A + B}{n - 2} (n U(x) - U_{1:n}(x) - U_{n:n}(x)) \]

\[ = \left( 1 - \frac{n}{n - 2} (A + B) \right) x + \left( A + \frac{A + B}{n - 2} \right) x^n + \left( B + \frac{A + B}{n - 2} \right) (1 - (1 - x)^n) \]

where the last step used the explicit formulas for \( U_{1:n} \) and \( U_{n:n} \). Comparing with the expression for \( \phi_w \) in (11), we find that the derived weighting function \( \phi_w \) matches (11) if we choose \( \varepsilon = \frac{n}{n-2} (A + B) \) and \( \varepsilon \theta = A + \frac{A + B}{n - 2} \) which implies \( \varepsilon (1 - \theta) = B + \frac{A + B}{n - 2} \). Solving for \( A \) and \( B \) and plugging into (22) shows that the learning rule and weighting function stated in the lemma are consistent. To complete the proof of the lemma, it now suffices to consider the second derivative of \( \phi''_w \) which is given by

\[ \phi''_w(x) = n(n - 1) \varepsilon \left( \theta x^{n-2} - (1 - \theta)(1 - x)^{n-2} \right). \]

For the assumed case \( n \geq 3, \varepsilon > 0 \) and \( \theta \in (0, 1) \), the terms outside the brackets are strictly positive. Moreover, the term in brackets is the difference between the strictly increasing positive function \( \theta x^{n-2} \) which starts in 0 for \( x = 0 \), and the strictly decreasing positive function \( (1 - \theta)(1 - x)^{n-2} \) which vanishes for \( x = 1 \). Thus, the term in brackets switches signs exactly once, from negative to positive. This confirms the inverse S-shape of \( \phi_w \). The inflection point is \( x_0 \) can be calculated directly from the condition \( \phi''_w(x_0) = 0 \).

Proof of Proposition 2. Notice first that we can write the distorted distribution function \( F_\varepsilon \) as a mixture of the original distribution \( F \) and the distributions functions of the extreme order statistics \( F_{1:n} \) and \( F_{n:n} \),

\[ F_\varepsilon(x) = \phi_w(F(x)) = (1 - \varepsilon) F(x) + \varepsilon F_{1:n}(x) + \varepsilon (1 - \theta) F_{n:n}(x). \]
Accordingly, the distorted mean can be written as the mixture

$E_\varepsilon[X] = (1 - \varepsilon)E[X] + \varepsilon\theta E[X_{1:n}] + \varepsilon(1 - \theta)E[X_{n:n}]$.

Rewriting this further as

$E_\varepsilon[X] = E[X] + \varepsilon \left( \theta E[X_{1:n}] + (1 - \theta)E[X_{n:n}] - E[X] \right)$

$= E[X] + \varepsilon \left( \left( \theta - \frac{1}{2} \right) (E[X_{1:n}] - E[X_{n:n}]) + \frac{1}{2} E[X_{1:n}] + \frac{1}{2} E[X_{n:n}] - E[X] \right)$

$= E[X] + \varepsilon \left( \theta - \frac{1}{2} \right) s_n(X) + \varepsilon s_n(X) \frac{\frac{1}{2} E[X_{1:n}] + \frac{1}{2} E[X_{n:n}] - E[X]}{E[X_{1:n}] - E[X_{n:n}]}$

leads to the representation (13). It remains to show that $s_n(X)$ is a valid measure of scale while $\gamma_n(X)$ is a valid measure of skewness. To demonstrate that $s_n(X)$ is a valid measure of scale, we rely on Definition 4.3 in Oja (1981), choosing for his order $\lesssim$ the dispersive order. $F$ is more dispersed than $G$ in the sense of the dispersive order, $G \lesssim_{disp} F$, if all pairs of quantiles lie further apart under $F$ than under $G$, $F^{-1}(u) - F^{-1}(v) \geq G^{-1}(u) - G^{-1}(v)$ for all $u, v \in (0, 1)$ with $u \geq v$. It follows from Theorem 3.B.31 in Shaked and Shanthikumar (2007) that $G \lesssim_{disp} F$ implies

$s_n(X) = E[X_{1:n}] - E[X_{n:n}] \geq E[Y_{1:n}] - E[Y_{n:n}] = s_n(Y)$

if $X \sim F$ and $Y \sim G$. The second property we need to verify is that $s_n(aX + b) = |a|s_n(X)$ for any real numbers $a$ and $b$. This follows immediately, once we notice that

$s_n(-X) = E[(-X)_{1:n}] - E[(-X)_{n:n}] = E[-X_{n:n}] - E[-X_{1:n}] = s_n(X)$.

To show that $\gamma_n(X)$ is a valid measure of skewness, we rely on definition 5.4 of Oja (1981). In line with the definition of skewness measures in Groeneveld and Meeden (1984), we choose for Oja’s order $\lesssim$ the convex transform order. We thus have to show first that if for two distribution functions $F$ and $G$ the function $R(x) = G^{-1}(F(x))$ is convex then $Y \sim G$ and $X \sim F$ implies $\gamma_n(Y) \geq \gamma_n(X)$.$^{17}$ To this end, let us focus first on the case where $F$ and $G$ are continuous with density functions $f$ and $g$. Notice that we can write $\gamma_n(X)$ via $F, F_{1:n}$

$^{17}$Our basic line of argument here follows Hosking (1989) who consider $L$-moments, which essentially (but not exactly) correspond to the special case $n = 3$. 

34
and \( F_{n:n} \) as
\[
\gamma_n(X) = \frac{\int_{-\infty}^{\infty} F(x) - \frac{1}{2} F_{1:n}(x) - \frac{1}{2} F_{n:n}(x) \, dx}{\int_{-\infty}^{\infty} F_{n:n}(x) - F_{1:n}(x) \, dx} \]
\[
= \frac{\int_{-\infty}^{\infty} h(F(x)) q(F(x)) \, dx}{\int_{-\infty}^{\infty} q(F(x)) \, dx}
\]

where
\[
h(x) = \frac{x - \frac{1}{2} U_{1:n}(x) - \frac{1}{2} U_{n:n}(x)}{U_{n:n}(x) - U_{1:n}(x)} = \frac{x - \frac{1}{2} (x^n + 1 - (1 - x^n))}{1 - (1 - x)^n - x^n}
\]

and
\[
q(x) = U_{n:n}(x) - U_{1:n}(x) = 1 - (1 - x)^n - x^n.
\]

Similarly, we can, of course, write
\[
\gamma_n(Y) = \frac{\int_{-\infty}^{\infty} h(G(y)) q(G(y)) \, dy}{\int_{-\infty}^{\infty} q(G(y)) \, dy}
\]

Making the substitution \( y = R(x) = G^{-1}(F(x)) \) in both integrals, we find that
\[
\gamma_n(Y) = \frac{\int_{-\infty}^{\infty} h(F(x)) q(F(x)) r(x) \, dx}{\int_{-\infty}^{\infty} q(F(x)) r(x) \, dx}
\]

where \( r(x) = f(x)/g(G^{-1}(F(x))) \) is the (weak) derivative of \( R \). To conclude the argument, we rely on Chebyshev’s integral inequality (Mitrinovic and Vasic (1970), p.40, Theorem 10). The inequality states that
\[
\gamma_n(Y) = \frac{\int_{-\infty}^{\infty} h(F(x)) q(F(x)) r(x) \, dx}{\int_{-\infty}^{\infty} q(F(x)) r(x) \, dx} \geq \frac{\int_{-\infty}^{\infty} h(F(x)) q(F(x)) \, dx}{\int_{-\infty}^{\infty} q(F(x)) \, dx} = \gamma_n(X)
\]

if \( h(F(x)) \) and \( r(x) \) are increasing while \( q(F(x)) \) is non-negative. By assumption, \( R \) is convex and thus \( r \) is increasing. To see that \( q \) is non-negative, note that for \( x \in [0, 1] \)
\[
(1 - x)^n + x^n \leq \sum_{i=0}^{n} \binom{n}{i} x^i (1 - x)^{n-i} = 1.
\]

It remains to show that \( h \) is increasing. To this end, notice that we can write
\[
h(x) = \frac{x - x^n}{1 - (1 - x)^n - x^n} - \frac{1}{2}.
\]
We need to show that the derivative $h'(x)$ is non-negative. This derivative is given by

$$h'(x) = \frac{nx^{n-1}(1-x)^{n-1} + 1 - (1-x)^n - nx(1-x)^{n-1} - nx^{n-1}(1-x) - x^n}{(1 - (1-x)^n - x^n)^2} \geq \frac{nx^{n-1}(1-x)^{n-1} + 1 - \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i}}{(1 - (1-x)^n - x^n)^2}$$

$$= \frac{nx^{n-1}(1-x)^{n-1}}{(1 - (1-x)^n - x^n)^2} \geq 0.$$ 

where the inequality used that $n \geq 3$. Discrete random variables can be approximated by continuous ones to arbitrary precision so the result also holds in the discrete case. To conclude the proof that $\gamma_n(X)$ is a valid measure of skewness, it thus remains to verify Oja’s second property, $\gamma_n(aX + b) = \text{sign}(a)\gamma_n(X)$ for any real numbers $a$ and $b$. This follows immediately.

\[\square\]

**Appendix B - Tables in support of the Baseline Results**

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>stdev</th>
<th>median</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_{1:6}</td>
<td>0.0128</td>
<td>0.0105</td>
<td>0.0103</td>
<td>-0.0032</td>
<td>0.1158</td>
</tr>
<tr>
<td>r_{2:6}</td>
<td>0.0063</td>
<td>0.0063</td>
<td>0.0051</td>
<td>-0.0184</td>
<td>0.0632</td>
</tr>
<tr>
<td>r_{3:6}</td>
<td>0.0021</td>
<td>0.0047</td>
<td>0.0017</td>
<td>-0.0308</td>
<td>0.0365</td>
</tr>
<tr>
<td>r_{4:6}</td>
<td>-0.0015</td>
<td>0.0049</td>
<td>-0.0005</td>
<td>-0.0385</td>
<td>0.0258</td>
</tr>
<tr>
<td>r_{5:6}</td>
<td>-0.0055</td>
<td>0.0069</td>
<td>-0.0036</td>
<td>-0.0612</td>
<td>0.0084</td>
</tr>
<tr>
<td>r_{6:6}</td>
<td>-0.0117</td>
<td>0.0115</td>
<td>-0.0090</td>
<td>-0.2047</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

*Table 6: This table shows the descriptive statistics of the sorted lagged returns, where lags from two days up to 7 days are used in the sorting.*

**Appendix C - Alternative Specifications**

Instead of using week-to-week changes of beliefs, in Table 7 and Table 8 we perform regressions using bi-weekly changes. This setting provides more intermediate returns such that the the number of ranks when sorting returns increases. A larger number of ranks provides a better
insight on the attention allocation. The bi-weekly changes are derived analogously to (1),

$$\Delta \text{bullish}_t = \frac{\text{bullish}_t - \text{bullish}_{t-2}}{\text{bullish}_{t-2}}, \quad \Delta \text{bearish}_t = \frac{\text{bullish}_t - \text{bullish}_{t-2}}{\text{bullish}_{t-2}}$$

(23)

Similar to the baseline analysis, we regress the changes in the measures of beliefs on the lagged returns. We include 20 lagged returns.

Comparing the results in Table 7 for all regressions, we conclude that returns are significantly influencing belief changes from the 2nd lag up to lags of 11 days. Therefore we proceed by sorting these lags and performing regressions of the beliefs changes on these sorted lagged returns. The results are shown in Table 8.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$Bullish</td>
<td>-1.13*</td>
<td>-1.89*</td>
<td>0.94</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>(1.81)</td>
<td>(-1.86)</td>
<td>(1.52)</td>
<td>(-1.61)</td>
</tr>
<tr>
<td>$rt-1$</td>
<td>1.96***</td>
<td>-2.30***</td>
<td>1.69***</td>
<td>-1.98***</td>
</tr>
<tr>
<td></td>
<td>(3.72)</td>
<td>(-2.81)</td>
<td>(3.10)</td>
<td>(-2.31)</td>
</tr>
<tr>
<td>$rt-2$</td>
<td>3.71***</td>
<td>-4.90***</td>
<td>3.49***</td>
<td>-4.58***</td>
</tr>
<tr>
<td></td>
<td>(6.93)</td>
<td>(-4.91)</td>
<td>(6.50)</td>
<td>(-4.63)</td>
</tr>
<tr>
<td>$rt-3$</td>
<td>2.75***</td>
<td>-5.30***</td>
<td>2.57***</td>
<td>-5.04***</td>
</tr>
<tr>
<td></td>
<td>(3.72)</td>
<td>(-6.00)</td>
<td>(3.50)</td>
<td>(-5.68)</td>
</tr>
<tr>
<td>$rt-4$</td>
<td>2.85***</td>
<td>-3.47***</td>
<td>2.65***</td>
<td>-3.18***</td>
</tr>
<tr>
<td></td>
<td>(4.69)</td>
<td>(-4.26)</td>
<td>(4.26)</td>
<td>(-3.98)</td>
</tr>
<tr>
<td>$rt-5$</td>
<td>3.63***</td>
<td>-4.11***</td>
<td>3.42***</td>
<td>-3.82***</td>
</tr>
<tr>
<td></td>
<td>(6.03)</td>
<td>(-4.53)</td>
<td>(5.55)</td>
<td>(-4.22)</td>
</tr>
<tr>
<td>$rt-6$</td>
<td>4.42***</td>
<td>-4.56***</td>
<td>4.22***</td>
<td>-4.27***</td>
</tr>
<tr>
<td></td>
<td>(7.44)</td>
<td>(5.17)</td>
<td>(6.98)</td>
<td>(4.84)</td>
</tr>
<tr>
<td>$rt-7$</td>
<td>2.49***</td>
<td>-4.11***</td>
<td>2.29***</td>
<td>-3.82***</td>
</tr>
<tr>
<td></td>
<td>(4.07)</td>
<td>(5.16)</td>
<td>(3.71)</td>
<td>(-7.40)</td>
</tr>
<tr>
<td>$rt-8$</td>
<td>2.98***</td>
<td>-4.43***</td>
<td>2.77***</td>
<td>-4.13***</td>
</tr>
<tr>
<td></td>
<td>(4.22)</td>
<td>(-3.67)</td>
<td>(3.98)</td>
<td>(-3.49)</td>
</tr>
<tr>
<td>$rt-9$</td>
<td>2.99***</td>
<td>-3.07***</td>
<td>2.81***</td>
<td>-2.81***</td>
</tr>
<tr>
<td></td>
<td>(4.43)</td>
<td>(-3.23)</td>
<td>(4.14)</td>
<td>(-2.92)</td>
</tr>
<tr>
<td>$rt-10$</td>
<td>2.15***</td>
<td>-2.73***</td>
<td>1.98***</td>
<td>-2.48***</td>
</tr>
<tr>
<td></td>
<td>(3.78)</td>
<td>(-3.23)</td>
<td>(3.47)</td>
<td>(-2.89)</td>
</tr>
<tr>
<td>$rt-11$</td>
<td>1.22**</td>
<td>-1.79**</td>
<td>1.03*</td>
<td>-1.52**</td>
</tr>
<tr>
<td></td>
<td>(2.00)</td>
<td>(-2.50)</td>
<td>(1.70)</td>
<td>(-2.09)</td>
</tr>
<tr>
<td>$rt-12$</td>
<td>-2.43***</td>
<td>0.97</td>
<td>-2.62***</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>(-4.08)</td>
<td>(1.21)</td>
<td>(-4.43)</td>
<td>(1.59)</td>
</tr>
<tr>
<td>$rt-13$</td>
<td>-0.98</td>
<td>1.33**</td>
<td>-1.15*</td>
<td>1.59***</td>
</tr>
<tr>
<td></td>
<td>(-1.56)</td>
<td>(2.26)</td>
<td>(1.83)</td>
<td>(2.63)</td>
</tr>
<tr>
<td>$rt-14$</td>
<td>-0.39</td>
<td>-0.68</td>
<td>-0.57</td>
<td>-0.43</td>
</tr>
<tr>
<td></td>
<td>(-0.68)</td>
<td>(-0.99)</td>
<td>(-1.00)</td>
<td>(-0.61)</td>
</tr>
<tr>
<td>$rt-15$</td>
<td>-2.19***</td>
<td>1.81**</td>
<td>-2.42***</td>
<td>2.15***</td>
</tr>
<tr>
<td></td>
<td>(-4.12)</td>
<td>(2.44)</td>
<td>(-4.61)</td>
<td>(2.78)</td>
</tr>
<tr>
<td>$rt-16$</td>
<td>-0.88*</td>
<td>1.84**</td>
<td>-1.11**</td>
<td>2.17***</td>
</tr>
<tr>
<td></td>
<td>(-1.75)</td>
<td>(2.39)</td>
<td>(-2.16)</td>
<td>(2.75)</td>
</tr>
<tr>
<td>$rt-17$</td>
<td>-2.32***</td>
<td>2.83***</td>
<td>-2.51***</td>
<td>3.11***</td>
</tr>
<tr>
<td></td>
<td>(-4.59)</td>
<td>(3.31)</td>
<td>(-5.07)</td>
<td>(3.55)</td>
</tr>
<tr>
<td>$rt-18$</td>
<td>-0.68</td>
<td>0.71</td>
<td>-0.89</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>(-1.01)</td>
<td>(0.95)</td>
<td>(-1.30)</td>
<td>(1.34)</td>
</tr>
<tr>
<td>$rt-19$</td>
<td>-1.52***</td>
<td>1.54**</td>
<td>-1.75***</td>
<td>1.87***</td>
</tr>
<tr>
<td></td>
<td>(-2.72)</td>
<td>(2.41)</td>
<td>(-3.06)</td>
<td>(2.86)</td>
</tr>
<tr>
<td>$rt-20$</td>
<td>-1.52***</td>
<td>1.54**</td>
<td>-1.75***</td>
<td>1.87***</td>
</tr>
<tr>
<td></td>
<td>(-2.72)</td>
<td>(2.41)</td>
<td>(-3.06)</td>
<td>(2.86)</td>
</tr>
</tbody>
</table>

| Constant | 0.02***     | 0.06***      | 0.01         | 0.08***      |
|          | (4.71)      | (8.21)       | (1.28)       | (7.40)       |
| pos-dum  | -           | -            | 0.03**       | -0.05***     |
|          |             |              | (2.30)       | (-2.69)      |

| N      | 1646        | 1646         | 1646         | 1646         |
| adj. $R^2$ | 0.1594    | 0.1172       | 0.1616       | 0.1194       |

Table 7: The results show the significance of regressions of bi-weekly changes is measures of beliefs (as defined in (23)) on lagged returns. Newey and West (1987)-standard errors are reported between brackets.
Table 8: The results show the significance of regressions of bi-weekly changes of belief measures (as defined in (23)) on sorted lagged returns. Newey and West (1987)-standard errors are reported between brackets.

The results in Table 8 show similar effects as our baseline regressions. The extreme returns gain most attention relative to the intermediate returns.

Next we perform some alternative specifications to the baseline sortings. Note that the following results are again on weekly basis (instead of the bi-weekly frequency used above). In Table 9 and 10 we perform sorted regressions using the most recent five and most recent seven daily lags respectively. Both specifications include the first (insignificant lag). Our finding based on these results is again that attention is directed towards extreme returns.
Table 9: The results show the significance of regressions of weekly changes of belief measures on sorted lagged returns. Lags include the first lag up to the fifth. Newey and West (1987)-standard errors are reported between brackets.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>∆Bullish</td>
<td>∆Bearish</td>
<td>∆Bullish</td>
<td>∆Bearish</td>
</tr>
<tr>
<td>$r_{1:5}$</td>
<td>3.48***</td>
<td>-3.64***</td>
<td>3.37***</td>
<td>-3.55***</td>
</tr>
<tr>
<td></td>
<td>(5.44)</td>
<td>(-3.87)</td>
<td>(5.26)</td>
<td>(-3.74)</td>
</tr>
<tr>
<td>$r_{2:5}$</td>
<td>3.03**</td>
<td>-2.64</td>
<td>3.00**</td>
<td>-2.61</td>
</tr>
<tr>
<td></td>
<td>(2.51)</td>
<td>(-1.47)</td>
<td>(2.53)</td>
<td>(-1.48)</td>
</tr>
<tr>
<td>$r_{3:5}$</td>
<td>2.03</td>
<td>-2.53</td>
<td>0.42</td>
<td>-1.21</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(-0.95)</td>
<td>(0.23)</td>
<td>(-0.46)</td>
</tr>
<tr>
<td>$r_{4:5}$</td>
<td>0.08</td>
<td>-2.29</td>
<td>-0.17</td>
<td>-2.08</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(-1.17)</td>
<td>(-0.16)</td>
<td>(-1.06)</td>
</tr>
<tr>
<td>$r_{5:5}$</td>
<td>2.84***</td>
<td>-3.40***</td>
<td>2.87***</td>
<td>-3.42***</td>
</tr>
<tr>
<td></td>
<td>(3.97)</td>
<td>(-3.13)</td>
<td>(4.13)</td>
<td>(-3.21)</td>
</tr>
</tbody>
</table>

Constant: -0.00 0.05*** -0.02* 0.06***

(1) (2) (3) (4)

adj. $R^2$: 0.0483 0.0354 0.0502 0.0356

N: 1647 1647 1647 1647

Table 10: The results show the significance of regressions of weekly changes of belief measures on sorted lagged returns. Lags include the first lag up to the seventh. Newey and West (1987)-standard errors are reported between brackets.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>∆Bullish</td>
<td>∆Bearish</td>
<td>∆Bullish</td>
<td>∆Bearish</td>
</tr>
<tr>
<td>$r_{1:7}$</td>
<td>3.70***</td>
<td>-3.92***</td>
<td>3.49***</td>
<td>-3.74***</td>
</tr>
<tr>
<td></td>
<td>(5.25)</td>
<td>(-3.52)</td>
<td>(4.94)</td>
<td>(-3.37)</td>
</tr>
<tr>
<td>$r_{2:7}$</td>
<td>3.59**</td>
<td>-4.02**</td>
<td>3.62**</td>
<td>-4.04**</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(-2.02)</td>
<td>(2.56)</td>
<td>(-2.07)</td>
</tr>
<tr>
<td>$r_{3:7}$</td>
<td>1.42</td>
<td>0.27</td>
<td>1.60</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td>(0.09)</td>
<td>(0.75)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>$r_{4:7}$</td>
<td>0.96</td>
<td>-3.16</td>
<td>-1.73</td>
<td>-0.86</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(-0.84)</td>
<td>(-0.58)</td>
<td>(-0.22)</td>
</tr>
<tr>
<td>$r_{5:7}$</td>
<td>1.83</td>
<td>-4.15</td>
<td>1.70</td>
<td>-4.04</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(-1.34)</td>
<td>(0.94)</td>
<td>(-1.32)</td>
</tr>
<tr>
<td>$r_{6:7}$</td>
<td>1.82</td>
<td>-2.66</td>
<td>1.60</td>
<td>-2.46</td>
</tr>
<tr>
<td></td>
<td>(1.34)</td>
<td>(-1.26)</td>
<td>(1.19)</td>
<td>(-1.19)</td>
</tr>
<tr>
<td>$r_{7:7}$</td>
<td>2.61***</td>
<td>-2.66**</td>
<td>2.62***</td>
<td>-2.68**</td>
</tr>
<tr>
<td></td>
<td>(3.22)</td>
<td>(-2.23)</td>
<td>(3.41)</td>
<td>(-2.33)</td>
</tr>
</tbody>
</table>

Constant: -0.01 0.06*** -0.03*** 0.08***

(1) (2) (3) (4)

adj. $R^2$: 0.0722 0.0521 0.0754 0.0529

N: 1647 1647 1647 1647