The Decrease in Confidence with Forecast Extremity

Doron Sonsino, Yefim Roth

Abstract: Three return forecasting experiments and a panel of more than 14,000 CFOs' forecasts of the S&P 500 annual return suggest that forecast confidence decreases as the forecasts diverge from zero, in the positive or negative direction. The decrease in confidence reflects in longer forecast intervals, weaker belief in the accuracy of the forecasts, and larger perceived volatility estimates. Assuming cumulative prospect theory, the increase in perceived volatility with forecast optimism is fast enough to fully offset the CFOs' response to more optimistic expectations in about 20% of the cross-sample comparisons. Permutation tests, more generally, confirm that the decrease in confidence significantly delays the response to optimistic forecasts. The decrease in confidence alleviates the underestimation of volatility in cases of optimistic forecasts, but even the optimistic CFOs underestimate the VIX volatility by more than 50%. A complementary empirical analysis reveals significant cross-sectional and time-series correlations between the absolute realized returns on the stocks composing the S&P 500 list and estimates of the stocks' contemporaneous volatility. The correlations emerge in five levels of analysis and separately show for positive and negative sub-sequences of the returns.

We thank Itzhak Ben-David, John Graham and Campbell Harvey for providing us access to the Duke University CFO's forecasts.

1. Introduction

Markowitz (1952) canonical model of portfolio selection stipulates that investors show desire for expected return while exhibiting aversion to the variance. Surveys and experiments indeed confirm that investors positively respond to increase in the expected return, while negatively reacting to increased perceived volatility (see Calvo-Pardo et al., 2021 recent discussion). The current paper examines the return forecasts of the participants in three published experiments and in a large panel of chief financial officers, showing that forecast confidence decreases with the extremity of the forecasts. Parametric estimates of the perceived volatility accordingly increase with forecast extremity. Confining to the CFOs' positive return forecasts, we utilize cumulative prospect theory (Tversky and Kahneman, 1992) to test if the increase in perceived volatility with forecast optimism is fast enough to cancel out or significantly delay the response to optimistic expectations.

Similarly to Graham and Harvey (2001), Glaser and Weber (2007), Ben-David et al. (2013) and others, we derive the parametric perceived volatility estimates from confidence intervals that the participants submit for future returns, but we innovatively assume beta distributions to directly account for the asymmetry of the intervals. The analysis then reveals that in the domain of positive forecasts, the prediction intervals get longer and the volatility estimates consistently increase, as the forecasts turn more optimistic. The directions reverse in the domain of negative forecasts, where the intervals' length and the respective volatility estimates increase as the return forecasts become more pessimistic. Overall, two experiments (1 and 2) and the analysis of more than 14,000 CFOs' short-term and long-term S&P 500 return forecasts expose positive significant correlations between the extremity of the forecasts, measured in terms of the deviation of the forecast from zero, and estimates of the perceived volatility of the target return. When the extremity of the expected return forecasts is measured relatively to historical benchmarks, such as the realized return in the recent past, or the median forecast (an estimate of the current consensus), the correlations decrease or vanish, suggesting that the relevant reference point for defining extremity is the zero return. The correlations between the point forecasts in absolute value and the respective perceived volatility estimates robustly show across the samples (e.g., within the waves of the CFOs' panel), and also reflect at the respondent level (across repeated participations of the same CFO).

In the domain of positive expected return forecasts, the increase in perceived volatility (σ) with expected return ($\mu>0$) may delay and even fully offset the response to optimistic expectations. If $\mu'>\mu$, but $\sigma'>\sigma$ in parallel, then the willingness to invest in the (μ',σ') asset may be smaller than the willingness to invest in the (μ,σ) asset, in spite of the more optimistic expectations for the former asset. Assuming cumulative prospect theory (CPT) preferences, with the canonical parameters of Tversky and Kahneman (1992) or the recent international-study parameters of L'Haridon and Vieider (2019), we argue that the increase in perceived volatility significantly delays the CFOs response to optimistic expectations. At the cross-section, the increase in volatility with the expected return is steep enough to fully offset the impact of more optimistic expectations in about 20% of the possible comparisons. Permutation tests moreover reveal that the responsiveness to optimistic expectations significantly increases when the μ and σ correlations are dissolved. Similar, albeit weaker in magnitude, results emerge in a parallel respondent-level analysis.

The increase in the length of confidence intervals with forecast extremity is consistently asymmetric. In the domain of positive expected return forecasts, the intervals are skewed to the left, and the skewness to the left intensifies as the forecasts get more positive. The directions reverse in the domain of negative expected return forecasts, with the intervals appearing more skewed to the right as the forecasts turn more negative. De Bondt (1993) *forecast hedging* theory, suggests that forecasters hedge their return forecasts in a contrarian style. The return distributions of optimistic forecasters are skewed to the left, while the distributions of pessimistic forecasters are skewed to the right (see also Du and Budescu, 2007; Grosshans and Zeisberger, 2018; Zhu et al. 2021). The analyses of the current paper additionally reveal that the contrarian hedging intensifies as the forecasts turn more extreme. The parametric volatility estimates derived assuming the beta distribution consistently increase with the extremity of the expected return forecasts ($|\mu|$), while the respective skewness estimates decrease with the signed return forecasts (μ). Regression analyses, however, reveal that the increase in forecast hedging with forecast extremity roughly accounts for less than 1/3 of the σ and $|\mu|$ correlations, so that the increase in perceived volatility with forecast extremity is more general than forecast hedging alone. De Bondt (1993; page 369) concluding discussion suggests that "*the mere fact that a stock goes up in price, increases its downturn potential. Thus, investors may become reluctant to buy more shares*". We formally test this conjecture using a large panel of CFOs' return forecasts and assuming CPT preferences, additionally showing that increase in perceived volatility with forecast extremity sustains when the effect of forecast hedging is neutralized.

Essentially, the $|\mu|$ and σ correlations are attributed to a decrease in confidence as the forecasts turn more extreme. In interval forecasting, the smaller confidence reflects in longer confidence intervals, and in derivation of parametric volatility estimates the longer intervals translate into larger perceived volatilities. Experiment 3 directly shows that the likelihood that MBA students assign to return falling within intervals of plus/minus δ from their point predictions significantly decreases with the extremity of their point forecasts. The likelihood assigned to forecasting errors smaller than δ accordingly decreases with forecast extremity. The discussions relate the decrease in confidence with forecast extremity to proportional or relative thinking (see Thaler, 1980; Tversky and Kahneman, 1981 pioneering discussions). The interval $[\mu-\delta,\mu+\delta]$ may appear sufficiently large for small predictions μ, while appearing rather small for large, positive or negative, predictions. We also argue that the *anchoring and adjustment* heuristic (Tversky and Kahneman, 1974), combined with proportional thinking, may technically explain the increase in the length of forecast intervals with forecast extremity.

Motivated by the behavioral correlations, we conduct an exploratory empirical analysis of the links between realized volatility and return extremity in historical series. The analysis reveals significant, cross-sectional and time-series, correlations between the absolute realized return on the S&P 500 stocks and estimates of the stocks' contemporaneous volatility. The positive correlations show in five levels of analysis (daily, monthly, quarterly, half-yearly, and yearly), and separately emerge for the positive and the negative return sub-sequences in all levels of analysis. Similarly to the behavioral results, the correlations disappear and even change sign when the absolute returns are replaced by the signed values. Measures based on absolute returns have been advanced as useful volatility estimates (Granger and Ding, 1995; Ederington and Guan, 2006), and some studies find that lagged absolute returns exhibit stronger predictive power than lagged volatilities for future volatility (Ghysels et al., 2006). We are not aware, however, of preceding studies documenting correlational patterns similar to what we report in the current ad hoc examination. We keep the empirical analysis concise, briefly arguing that beyond the behavioral explanations, the $|\mu|$ and σ correlations documented in the main sections of the paper may have empirical roots.

At the level of interpretation, the decrease in confidence with forecast extremity connects with studies suggesting that private investors tend for contrarian trading (e.g., Grinblatt and Keloharju, 2000; Baltzer et al., 2019) and institutional investors benefit from momentum or feedback trading on the expense of private investors (Grinblat and Han, 2005; Barber et al., 2009; Economou et al., 2022). The results also relate to the literature on overconfidence and trading (Glaser and Weber, 2007; Ben-David et al., 2013). Using the panel data we show, for example, that the decrease in confidence with forecast extremity partially alleviates the underestimation of volatility in the sub-sample of relatively optimistic CFOs, and increases their willingness to protect against the S&P 500 volatility. However, the optimistic CFOs' forecast intervals are still about 50% shorter than implied by the concurrent VIX levels.

While the main result of the paper suggests that forecast uncertainty increases as the conditional return forecasts turn more optimistic, the CFOs' panel analysis also brings a secondary result linking to the literature on misperception of the risk-return tradeoff. The belief that riskier assets should compensate risk-averse investors in terms of higher expected returns is considered a fundamental premise of finance (https://www.nasdaq.com/glossary/r/risk-return-trade-off). Behavioral studies, however, demonstrate that investors stereotype companies and stocks, so that "good companies" or "good stocks" are considered safe investments that yield relatively high returns (see Shefrin and Statman, 1995; Ganzach, 2000; Shefrin, 2001 early evidence). At the higher level of macroeconomic expectations, positive expectations may increase optimism regarding the stock market prospects, while decreasing the perceived riskiness or volatility of stock investment (Amromin and Sharpe, 2014). Indeed, when we average the return forecasts and the respective perceived volatility estimates in each of the 50 quarters of the CFOs' panel, we find negative correlations between the mean (signed) expected returns and the mean perceived volatility estimates. Quarters with relatively optimistic expectations ($\bar{\mu}$) exhibit lower perceived volatility estimates ($\bar{\sigma}$) and vice versa. While the (short-term) $\bar{\mu}$ decreases with the VIX for the date of the survey and increases with the Michigan consumer sentiment index (SENT) for the survey date, $\bar{\sigma}$ conversely increases with VIX and decreases with SENT. Across the 50 quarters of the panel, our results match the findings in the literature on misperception of the risk-return tradeoff.

The paper proceeds as follows. Section 2 presents the evidence obtained in analyzing the links between forecast extremity and forecast confidence in three published experiments. For brevity, we discuss the method of the experiments and the sample very briefly, leaving the details for the Web supplements. The baseline hypothesis of the paper, asserting that "forecast confidence decreases with forecast extremity" is addressed as H1 for convenience. Section 3 proceeds to discuss the results for the panel of CFOs' forecasts. Section 4 explains how we utilize cumulative prospect theory to proxy the CFOs willingness to invest in the S&P 500, and presents the analysis testing the implications of H1 assuming CPT preferences. Section 5 reports the results of our empirical analysis of the absolute return and volatility correlations. Section 6 concludes, discussing the results and possible implications.

2. Experimental evidence

2.1: Experiment 1 – confidence intervals for future returns

Sonsino and Regev (2013) report the results of an email survey-experiment where individually-tailored questionnaires were distributed to a closed list of preregistered MBA students, alumni, and members of financial Web forums. The participants were presented with ten binary choice problems between familiar Israeli stocks, and provided 95% lower and upper confidence limits for the three months return on their selected stocks. The stocks for each binary choice problem were randomly drawn from the list of 25 largest stocks in the Tel-Aviv exchange, to avoid the bias the specific task selection may induce (e.g., Juslin, 1994). The instructions clarified that the interval between the lower and upper 95% confidence limits denotes a 90% confidence interval for the return, and explained the tradeoffs between submitting narrow intervals that might miss the actual return and excessively wide intervals that do not open possibility for 5% error. A bonus of 80 NIS (about 25 USD) was promised to participants that meet three conditions for fair and accurate interval prediction. Using P95 and P5 to denote the upper and lower confidence limits, the variable LGTH=P95-P5 henceforth represents the length of the prediction interval. Intuitively, longer intervals point at lower forecast confidence, so LGTH is henceforth used as an instant measure of forecast confidence. To formally derive a perceived volatility estimate from the interval, we use the Pearson and Tukey approximation that is considered a relatively accurate method for extracting the standard deviation from the 5% quantiles (Murphy and Winkler, 1974). We accordingly let $\sigma = \mathrm{LGTH}/3.25$ represent the perceived volatility of the target return.[1] The midpoint of the interval, m=(P95+P5)/2, is used to measure the distance of the interval from the origin, and the absolute value of the midpoint |m| is adopted as the measure of forecast extremity. The next paragraphs summarize the results of testing the |m| and LGTH (or σ) correlations, across the sample and at the participant level. The correlations are reported for LGTH, noting that the results for σ are identical. The sample consists of 93 participants, with 54% holding or pursuing an MBA and 40% reporting at least one year of professional investment-industry occupation. More details on the questionnaire design, the method of the experiment, and descriptive statistics for the main variables are presented in Web supplement A.

For the cross-sample analysis we average the data at the individual level, using upper bars to represent the participant-level means. The symbol $\overline{\mathrm{LGTH}}$ denotes the average length of the ten prediction intervals, and $\overline{|\mathrm{m}|}$ ($\overline{\mathrm{m}}$) are accordingly defined. The Pearson correlation

[1] Since the questionnaire did not elicit point predictions, the asymmetry of the intervals cannot be tested directly. Assuming normality, σ=LGTH/3.29, but since forecast intervals are rarely symmetric, the Pearson and Tukey approximation is preferred. The asymmetry is directly taken into account in the following sections.

between $\overline{|m|}$ and $\overline{\text{LGTH}}$ is 0.55, and a permutation test rejects equality to zero at p<0.01.[2] The correlation appears robust. It decreases to 0.36 (N=75; p<0.01) when the participants with the 10% most extreme average forecasts ($\overline{|m|}$) are removed, but increases to 0.49 (N=47; p<0.01) when the participants with the 25% most extreme average forecasts are ignored. About 82% of the N=930 prediction intervals submitted along the experiment have non-negative midpoints, so that |m|=m for most of the intervals. However, the correlation between $\overline{m}$ and $\overline{\text{LGTH}}$ is smaller 0.47, and the Hotelling Williams test for the equality of dependent correlations rejects equality of the ($\overline{|m|}$ and $\overline{\text{LGTH}}$) and ($\overline{m}$ and $\overline{\text{LGTH}}$) correlations at p≤0.05. By way of interpretation, the hypothesis that confidence decreases with optimism is rejected for a decrease in confidence with forecast extremity. The increase in lengths with extremity indeed shows "both ways", when the midpoints of the intervals turn more positive and when the midpoints turn more negative. The $\rho(\overline{|m|}, \overline{\text{LGTH}})$ is 0.66 based on the intervals with m>0, and 0.33 based on the smaller samples of intervals with m<0. A median split of the sample by the extremity of the forecasts reveals mean $\overline{\text{LGTH}}$ of 9.4 for the less extreme participants compared to more than half larger 14.5 for the more extreme participants (p<0.01). The respective mean σ estimates are 2.9 and 4.5.

The correlations also show at the individual level, despite the small samples of ten intervals per participant. When the correlation between $|m|$ and LGTH is calculated for each participant, the mean correlation is 0.41. The correlations are positive for 64 participants and negative for only 20, so a sign test suggests significance at p<0.01. When the absolute value sign is removed, testing the correlation between the signed midpoints (m) and LGTH, the mean correlation drops to 0.22, and paired comparisons show that the (m and LGTH) correlation exceeds the (|m| and LGTH) correlation for only six participants. A split of the ten prediction intervals of each participant by the median |m|, reveals a mean $\overline{\text{LGTH}}$ of 9.7 for the less extreme intervals compared to 14.1 for the more extreme intervals (p<0.01). The positive correlations show for all major sub-samples. The mean correlation is 0.45 for the MBAs (N=45; p<0.01), 0.42 for the participants with financial-industry experience (N=35; p<0.01), and 0.43 for the participants classified as risk-seeking in a subsidiary task

[2] Significance levels are two-tailed throughout the paper. We always report Pearson correlations, but the Spearman correlations are typically similar. The customary ρ is sometimes used for the correlations.

(N=32; p<0.01). Panel regressions moreover reveal that the extremity of the interval (measured by |m|) has twice stronger effect on the length of the intervals compared to diverse controls (results are provided in Web supplement A).

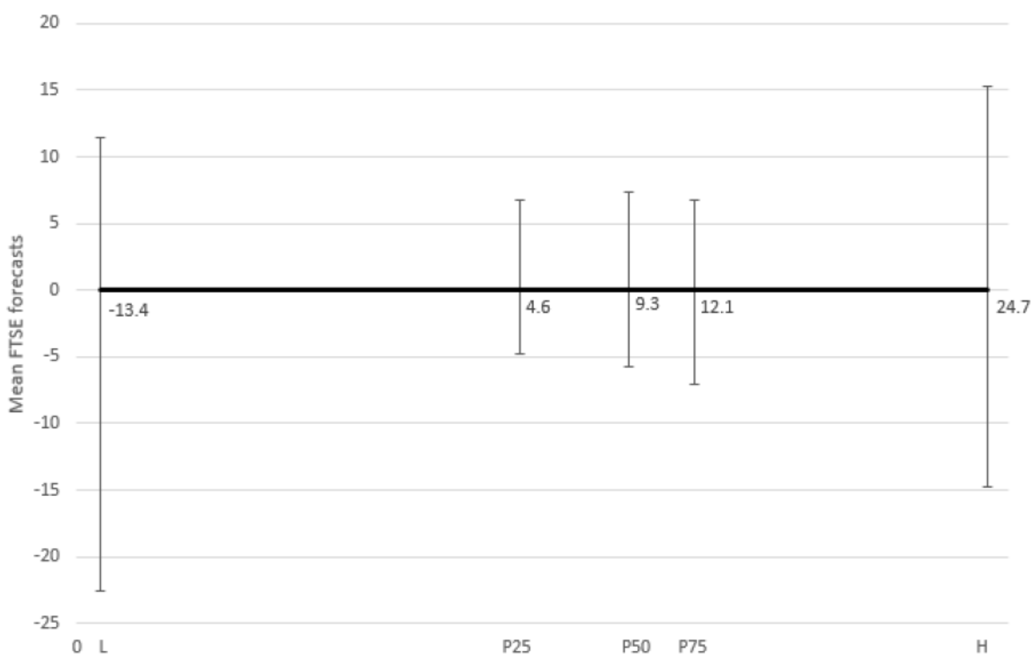
2.2: Experiment 2 – quartile forecasts for FTSE's return

Sonsino et al. (2022) report the results of a lab experiment using the exchangeability method (Baillon, 2008) to elicit probabilistic return forecasts from advanced business and economics students. The prediction target was the return on the *FTSE all shares* index (henceforth addressed as FTSE), over a period of twelve months following the lab sessions. To initiate the elicitations, the exchangeability program asks for lower (L) and upper (H) bounds, representing the most extreme values that the return can take over the prediction period.[3] The algorithm then generates a sequence of binary choice problems, between a deposit that pays 5% yearly return when FTSE exceeds a given threshold, and a deposit that pays the 5% when FTSE falls below this threshold. The threshold is sequentially updated depending on choices, up to convergence to a level that makes the participant approximately indifferent between the two deposits. As the convergence threshold splits the [L,H] space of uncertainty into exchangeable or equiprobable events, it represents a median forecast P50 for the index return. The iterated choice method is then applied to elicit a lower quartile forecast P25 that splits the [L,P50] interval into equiprobable events, and an upper quartile forecast P75 that splits [P50,H] into equiprobable events. The final output of the elicitations is five forecasts L,P25,P50,P75,H that summarize the expectations of each participant regarding the FTSE performance over the twelve months investment period. The N=73 participants were mostly MBA or MA in economics students and advanced accounting students preparing to the certification exams. Beyond a fixed participation fee, a late bonus of up to 106 NIS (about 30 USD) was derived from one of the binary choices made in the experiment and paid at the end of the investment period. A detailed example for the elicitations and more details on the experiment are provided in Web supplement B.

[3] For more applications of the exchangeability method see Abdellaoui et al. (2011), Menapace et al. (2015), and Jiao (2020).

Figure 1 presents the mean elicited forecasts, illustrating that the participants were generally optimistic regarding the FTSE index performance over the investment period. Almost 95% of the elicited P50 were positive, with only four of the 73 participants converging to a negative P50. The distribution in Figure 1 appears skewed to the left, as the P50-L distance is almost 50% larger than H-P50, and P50-P25 is about 2/3 larger than P75-P50. Since 95% of the P50 forecasts are positive, the skewness to the left fits De Bondt's (1993) forecast hedging theory. At the individual level, 75% of the 69 participants with positive P50 exhibit (P50-P25)>(P75-P50) in line with Bowley's definition of skewness to the left (Groeneveld and Meeden, 1984), while all four participants with negative P50 conversely show P75-P50>P50-P25.

<u>Figure 1: The FTSE Forecasts</u>



<u>Note:</u> The bars represent the plus and minus 40% range around the means.

Noting that [P25,P75] constitutes a 50% confidence interval for the FTSE return, we first test H1 using the correlations between |P50| and P75-P25. As in experiment 1, longer intervals propose smaller forecast certainty, so the decrease in confidence with forecast extremity should reflect in increased P75-P25 distances as P50 diverges from zero.[4] Indeed, ρ(|P50|,P75-P25) is positive 0.32 ($p<0.01$), and it is robust to removing the participants with the 10% smallest or largest |P50| (see the leftmost column of Table I).[5] The mean P75-P25 distances increase from 5.4 when |P50|≤5, to 7.2 when 5<|P50|≤10, reaching 8.5 when |P50|>10. The correlation separately shows for the forecasts around positive P50, and for the anecdotal (N=4) sample of forecasts around negative P50. Moreover, the skewness to the left of the intervals around positive P50 appears to increase as P50 diverges from zero, as P50-P25 significantly increases while P75-P50 does not change consistently with P50 (see the "P50>0" row of Table I). The directions reverse for the anecdotal sample of intervals around negative forecasts, where P75-P25 increases with |P50| while P50-P25 tends to decrease with |P50| (see the "P50<0" row of the table). The correlational analysis thus suggests that forecast hedging intensifies as the forecasts turn more extreme.

Table I: The |P50| correlations in experiment 2

	P75-P25	P75-P50	P50-P25	σ	Skew
Full sample (N=73)	0.32***	0.16	0.29***	0.40***	-0.36***
Filtered sample (N=59) (10% extreme |P50| removed)	0.37***	-0.01	0.44***	0.32***	-0.40***
The P50>0 sample (N=69)	0.29**	-0.07	0.37***	0.34***	-0.50***
The P50<0 sample (N=4)	0.69	0.95*	-0.38	0.83**	0.99**

Notes: The left panel presents the correlations between the absolute median predictions |P50| and the distances P75-P25, P75-P50, P50-P25. The right panel presents the |P50| correlations with the parametric σ and Skew estimates derived as explained in the text. The upmost row reports the correlations for the full sample (N=73). The "filtered sample" row presents the correlations after removing the 7 participants with the largest and smallest |P50| (N=59). The two rows at the bottom present the correlations for the intervals around positive P50 (N=69) and for the small sample of intervals around negative P50 (N=4). The asterisks summarize the results of permutation tests on the correlations, with one/two/three asterisks representing significance at p≤0.01/0.05/0.1.

[4] Note that P75-P25 serves as a measure of forecast confidence independently of the bounds L and H. Consider, for example, the case where the exchangeability program presents a binary choice between deposit A that pay 5% return when FTSE falls within [L,P50-4%] and deposit B that pays the 5% when FTSE falls within [P50-4%,P50]. If the preference for deposit A over deposit B increases with |P50|, then the likelihood that the participants assign to a 4% negative prediction error decreases and the P50-P25 distance tend to increases with |P50|. A similar argument applies in the opposite direction, suggesting that an increase in P75-P50 with |P50| is a signal for lower confidence in P50, independently of H.

[5] The correlation between |P50| and (H-L) is about 50% smaller, 0.15. We conjecture that the weak correlation is an artefact of the method and if P50 was elicited before the bounds, the correlation would have increased.

To test H1 more generally, building on the five elicited forecasts and taking the skewness into account, we assume the target return is beta distributed on the interval [L,H]. The beta distribution is commonly used to fit elicited beliefs and forecasts in various domains (e.g., Abbas et el., 2008; Engelberg et al., 2009; Manski and Neri, 2013; Wallsten et al., 2016; Eyting and Schmidt, 2021). The four-parameters beta, henceforth denoted Beta(*a*,*b*,L,H), has shape parameters *a* and *b*, with L and H representing the support of the distribution. Depending on the shape parameters, the distribution can fit negatively skewed or positively skewed distributions, as well as symmetric distributions (Johnson et al., 1995). We apply non-linear OLS to estimate the shape parameters *a* and *b* from the three quartiles P25, P50, and P75. The estimations robustly converge, with the predicted quartiles showing close to perfect correlations between the elicited quartiles. The mean absolute error, defined as the distance between the predicted quartiles and the elicited quartiles, is 0.63. The mean normalized error, derived by dividing the absolute error by H-L, is 1.8%.[6] Parametric perceived volatility and skewness estimates are derived from the estimated beta parameters using the standard formulas:

$$(1) \quad \sigma = (H - L)\sqrt{\frac{ab}{(a+b)^2\,(a+b+1)}}$$

$$(2) \quad Skew = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$$

The mean estimated σ following this approach is 5.6 with a standard deviation (STD) of 3.5. The mean Skew is negative -0.24 (STD 0.59). The correlations between |P50| and the parametric σ and Skew estimates are disclosed at the right columns of Table I. The |P50| and σ correlation is 0.40 (p<0.01), with the mean σ climbing from 3.9 when |P50|≤5, to 5.4 when 5<|P50|≤10, reaching 6.5 when |P50|>10. The |P50| correlation with Skew is negative -0.50 (p<0.01) for the 69 participants with P50>0, and positive 0.99 for the four participants

[6] By the definition of the four parameters beta distribution, if X is beta(*a*,*b*,L,H), the normalized variable (X-L)/(H-L) is beta distributed on [0,1] with the same shape parameters, which we denote Beta(*a*,*b*). The shape parameters can thus be estimated from the three equations (P50-L)/(H-L)=the 0.5 quantile of Beta(*a*,*b*), (P75-L)/(H-L)=the 0.75 quantile of Beta(*a*,*b*), (P25-L)/(H-L)=the 0.25 quantile of Beta(*a*,*b*). The correlations between the elicited quantiles and the predicted quantiles are 0.99 for P50 and P75 and 0.98 for P25. The mean normalized prediction errors are 1.5%, 2.1% and 1.9%, respectively.

with P50<0. The estimated skewness accordingly decreases with P50 across the sample, with ρ(P50,Skew)= -0.56. Regressions controlling for personal attributes of the participants including socio-demographics, the big-five personality traits, and a measure of personal risk tolerance, suggest that |P50| is the single variable that robustly and significantly affects the P75-P25 distances and the estimated σ parameters, while P50 is the only variable that robustly affects the skewness estimates (see Web supplement B).

Finally, note that the correlations of Table I drop when the extremity of the forecasts is defined relatively to FTSE's performance in the recent past as presented in a handout that was distributed with the printed instructions. The most recent annual FTSE return, as shown in the handout, was 16.8%. If the recent FTSE return is the relevant anchor for measuring the extremity of the forecasts, then the forecast extremity measure should be changed to |P50-16.8|. However, the correlations between |P50-16.8| and P75-P25 (ρ=0.12) or σ (ρ=0.20) are more than half smaller than the |P50| correlations, and the correlations similarly decline when using other historical benchmarks, or the current mean or median P50 forecasts (as estimates of the consensus).[7]

2.3: Experiment 3 – forecast confidence statements

Sonsino et al. (2021) explore a modified interval forecasting task, where the respondents submit a median prediction for the quarterly return on a familiar stock, and assess the likelihood of the return falling within a distance of δ from their prediction. When the median prediction is 2% and δ=5%, for example, the participant estimates the likelihood of return falling within the [-3%,7%] interval. Essentially, the respondents assess the likelihood of exhibiting a prediction error smaller than δ, but the task is neutrally framed, avoiding terms such as error or accuracy. The printed questionnaire consisted of eight tasks, with each task referring to a given stock from the Tel-Aviv exchange. The margin δ took values of 5% or 10%, and participation was incentivized using binarized scoring rules (Hossain and Okui, 2013). The questionnaire was distributed in MBA finance classes, and the sample consists of N=72 MBAs. The variable CONF, taking values between 0% and

[7] Moreover, the correlations between P50 and P75-P25 (σ) based on the N=32 forecasts with 0<P50<9.53 (the median) are positive 0.42 (0.38).

100%, is henceforth used for the likelihood assigned to the [P50-δ, P50+δ] interval, with P50 representing the direct (median) point forecast. More details on the questionnaire, the incentivization, and descriptive statistics for the main variables are provided in Web supplement C. Currently, we note that the forecast and the confidence variables are highly clustered, with |P50| taking 15-22 values and the CONF statements taking 13-16 values in each of the eight tasks. The P50 forecasts are 83% positive, taking negative values in only 14% of the cases. Hypothesis H1, in the context of experiment 3, predicts that CONF should decrease with |P50|, as the participants' confidence in exhibiting a prediction error smaller than given δ decreases as their forecasts turn more extreme.

Table II: The task-level |P50| and CONF correlations

	T1	T2	T3	T4	T5	T6	T7	T8	Aggregate
ρ1	-0.06	-0.13	-0.12	-0.27**	-0.21*	-0.23**	-0.16	0.07	-0.33***
ρ2	-0.08 (N=21)	-0.20 (N=17)	-0.43* (N=20)	-0.52*** (N=22)	-0.36 (N=16)	-0.38* (N=19)	-0.49** (N=16)	0.16 (N=15)	-0.65** (N=12)

Notes: T1-T8 represent the eight forecasting tasks, and the left columns of the table present the task-level correlations between |P50| and CONF. ρ1 is the correlation based on the full sample of 72 participants. For ρ2, the CONF values are averaged for each |P50| before deriving the correlation. The number of observations is presented in smaller font parentheses. The "Aggregate" column at the right of the table presents the correlations between measures of *forecast-extremity* and *forecast-confidence* based on all eight tasks. Forecast-extremity is derived by normalizing the |P50| in each task to zero mean and unit variance, and averaging the eight normalized |P50|. Forecast-confidence is measured similarly. The "Aggregate" column ρ1 presents the correlation between forecast-extremity and forecast-confidence for the full sample of 72 participants. For ρ2, the sample is sorted by forecast-extremity, and grouped into twelve bins consisting of six participants with similar forecast-extremity scores. The table presents the correlation between the mean forecast-extremity and the mean forecast-confidence measures for the twelve groups. The asterisks summarize the results of permutation tests on the correlations, with one/two/three asterisks representing significance at $p \leq 0.01/0.05/0.1$.

To start the analysis, we examine the task-level results. The top row of Table II presents the |P50 | and CONF correlations in each task, based on the full sample of 72 participants. The correlations are negative but weak, ranging between -0.06 and -0.27, in seven of the eight cases. However, the correlations get stronger when the cluster structure of the data is utilized to decrease noise. The second row of the table presents the correlation between the average CONF, for each level of |P50|, and the respective |P50|. The correlations are negative, ranging between -0.20 and -0.52 in six of the eight tasks. Overall, the task-specific results support H1, but the evidence is weak, possibly because of the idiosyncratic noise in the task-level responses. Indeed, when the |P50| and CONF in each task are normalized to have zero mean and unit variance, and the eight normalized variables are averaged to obtain individual *forecast-extremity* and *forecast-confidence* scores, the correlation between the two scores is -0.33 ($p<0.01$).[8] When the sample is sorted by forecast-extremity and grouped into 12 bins of six participants with similar forecast-extremity scores, the correlation between the mean forecast-extremity and mean forecast-confidence scores (across the 12 bins) is -0.65 ($p=0.02$). Returning to the non-normalized forecast confidence scores, a split into quartiles by forecast-extremity reveals mean CONF 82 for the participants in the least extreme quartile, compared to 67 for the participants in the most extreme quartile (N=18 in each sample; $p<0.01$). The participants in the intermediate forecast-extremity quartiles have a similar mean CONF of about 75.

Panel regressions additionally confirm that confidence still decreases with forecast extremity when socio-demographics and personality characteristics are controlled. The only variable that robustly affects CONF beyond |P50| is the personality trait of neuroticism. The cross-sample correlation between neuroticism and the individual forecast-confidence scores is -0.23 ($p=0.05$), compared to the -0.33 ($p<0.01$) forecast-extremity and forecast-confidence correlation. Neuroticism does not correlate with forecast-extremity ($\rho=-0.03$) and the panel regressions show that CONF jointly decreases with |P50| and neuroticism (see Web supplement C for results). The mean CONF of the respondents that rank above median in forecast-extremity and also rank above median in neuroticism is 66, compared to 85 for the respondents that rank below median in both variables ($p<0.01$).

3. The panel of CFOs' forecasts

3.1: The data and notation

This section turns to testing the links between forecast extremity and perceived volatility in a panel of finance professionals' S&P 500 return forecasts. The panel survey was run by Duke University on a quarterly basis and the respondents are senior finance executives,

[8] The results for task 8 appear different from those for the other tasks: the P50 forecasts are positive in only 39% of the cases, compared to 72%-99% positive forecasts in the other tasks, and the task 8 CONF is 8 percentile points smaller, on average, from the CONF in the other tasks with the same δ. When the forecast-extremity and forecast-confidence scores are recalculated ignoring task 8, the correlation is -0.36 ($p<0.01$).

mostly CFOs.[9] We analyze the 14,168 forecasts provided in the 50 quarters between 2005:Q2 and 2017:Q3. The number of observations per quarter ranges between 174 and 417, with a mean close to 283. The number of respondents is 2,740, with between 2 and 44 observations per participant. The data consists of expected return forecasts and 80% confidence intervals for the S&P 500 (henceforth SP500) return in the short-term (ST) and the long-term (LT). The short-term question is presented in Figure 2.

Figure 2: The Duke panel short-term forecasting question

Over the next year, I expect the annual S&P 500 return will be:

- There is a 1-in-10 chance the actual return will be less than ___%
- I expect the return to be: ___%
- There is a 1-in-10 chance the actual return will be greater than ___%

The long-term forecasting question was identical, except for referring to the mean annual S&P 500 return over the next ten years.

The variable μ is henceforth used for the expected-return forecasts, with LGTH=P90-P10 denoting the distance between the upper (P90) and lower (P10) confidence limits. Ordered pairs (μ; [P10,P90]) are addressed as "*observations*" or "*forecasts*" from the panel, and terms such as "*the intervals around X%*" are used for the observations with μ=X%. The "*left (right) margin*" of the forecast interval is [P10,μ] ([μ,P90]).

Table III presents descriptive statistics for the short-term and long-term forecasts. The mean expected return forecast is 5% for the short-term and 6.8% for the long-term. The ST return forecasts are 88% positive, while the LT forecasts are almost always (98.8%) positive. The LT intervals are about 25% shorter than the ST intervals, suggesting that the respondents partially acknowledge the lower volatility of the ten-years average return. In addition, the intervals around positive return forecasts are more than 1/3 shorter than the intervals around negative or zero forecasts.

[9] Currently, the survey is run in partnership with the Federal Reserve Banks of Richmond and Atlanta (https://www.richmondfed.org/cfosurvey). For more discussions of the survey see Ben-David et al. (2013), Boutros et al. (2021) and the references therein.

Table III: The panel data

	ST forecasts	LT forecasts
Expected return (**μ**)	5.0 (5.1)	6.8 (4.8)
Proportion of μ > 0	87.9%	98.8%
Proportion of μ = 0	6.7%	0.8%
Proportion of μ < 0	5.4%	0.4%
Low confidence limit (**P10**)	-3.8 (9.4)	1.1 (6.0)
High confidence limit (**P90**)	10.6 (6.8)	11.0 (7.9)
Length of the intervals (**LGTH**)	14.5 (12.0)	9.9 (9.4)
Length of the intervals around μ > 0	13.3 (11.2)	9.8 (9.2)
Length of the intervals around μ = 0	20.7 (12.8)	19.4 (17.1)
Length of the intervals around μ < 0	25.8 (13.9)	25.3 (13.9)
Perceived volatility estimate (**σ**)	5.5 (4.6)	3.8 (3.6)
Skewness estimate (**skew**)	-0.28 (0.6)	-0.21 (0.6)
Skew of intervals around μ > 0	-0.29 (0.6)	-0.22 (0.6)
Skew of intervals around μ < 0	-0.01 (0.5)	-0.12 (0.6)
Skew of intervals around μ < -5%	0.11 (0.5)	-0.05 (0.6)

Notes: The table presents the means, with the standard deviation in smaller brackets, or the proportions for (subsets of) the 14,168 observations in the panel. The percentage signs are suppressed when presenting the forecast-related variables.

About 79% of the ST intervals and 77% of the LT intervals are asymmetric, with μ-P10 ≠ P90-μ. In line with De Bondt's (1993) forecast hedging, the intervals around positive μ appear skewed to the left, with μ-P10>P90-μ for 71% of the asymmetric intervals. The intervals around negative μ are mixed, with P90-μ>μ-P10 for 47% of the asymmetric intervals, but the proportion increases to 60% for the intervals around μ<-5%.[10] To derive parametric volatility and skewness estimates, we assume that the return is beta distributed on an interval extending the [P10,P90] confidence interval. Following a test of several alternatives, we chose to extend the left and right margins by 25%, assuming L=P10-0.25·(μ-P10) and H=P90+0.25·(P90-μ). As in experiment 3, the estimations fit the data well, with mean relative prediction errors of 2.2% for ST and 2.0% for LT.[11] The mean

[10] The weaker skewness to the right in cases of negative expectations, compared to the definite skewness to the left in cases of positive expectations, also shows in De Bondt (1993). The parametric skewness estimates are always negative when (μ-P10)>(P90-μ) and positive when (μ-P10)<(P90-μ).

[11] The three estimated equations, after substituting the L and H and rearranging, are (μ-P10)/(P90-P10) = expected value of Beta(*a*,*b*); [(P90-P10)+0.25·(μ-P10)] / [1.25·(P90-P10)] = 90% quantile of Beta(*a*,*b*); [0.25·(μ-P10)] / [1.25·(P90-P10)] = 10% quantile of Beta(*a*,*b*). The parameters *a* and *b* therefore only depend

parametric σ estimates are 5.5% for the ST forecasts and 3.8% for the LT forecasts. The mean skewness estimates are negative for the intervals around positive μ, and positive for the ST intervals around μ<-5%.

3.2: A preliminary glimpse at the forecast extremity effects

Table IV reports the results of comparing the length and the estimated σ and Skew of the more extreme and less extreme forecasts (the panel structure of the data is ignored for the brief glimpse, but taken into account in the following sections). The upmost panel shows that the intervals around more extreme forecasts are longer, and the increase in length shows for the ST and LT intervals, and for the intervals around positive and negative μ. The intermediate panel similarly shows that the σ estimates increase with the extremity of the forecasts in all four comparisons. The panel at the bottom then shows that the intervals around positive μ are negatively skewed, and the skewness is stronger for the extreme forecasts. The directions reverse for the smaller samples of intervals around a negative μ, where the skewness increases with |μ|.

Table IV: Median splits by the extremity of the forecasts

	ST forecasts		LT forecasts	
	Positive μ	**Negative μ**	**Positive μ**	**Negative μ**
LGTH				
|μ| < |median|	9.8	15.1	8.2	14.1
|μ| > |median|	15.7	33.0	11.4	34.6
σ				
|μ| < |median|	3.8	5.7	3.1	5.3
|μ| > |median|	6.0	12.4	4.3	13.2
Skew				
|μ| < |median|	-0.24	-0.08	-0.16	-0.22
|μ| > |median|	-0.33	0.11	-0.25	-0.05

Notes: The positive and negative, ST and LT forecasts are median split by |μ| and the table presents the mean LGTH, σ and Skew for the |μ| < |median| and |μ| > |median| observations. The percentage signs are suppressed in presenting the LGTH, σ, and Skew variables. The number of observations in each cell and more details are provided in an extended Web supplement D version of the table.

on the distances μ-P10 and P90-μ; i.e., the estimated *a* and *b* for the observations (μ; [P10,P90]) and (μ'; [P10',P90']) are identical when μ-P10=μ'-P10' and P90-μ=P90'-μ'.

3.3: The increase in forecast hedging with forecast extremity

The results of experiment 3 suggested that De Bondt's (1993) forecast hedging intensifies with the extremity of the forecasts, but the sample was small (N=72, with only four negative forecasts). To test if the pattern shows in the panel of SP500 forecasts, we estimate the equations

$$(3) \qquad (\mu - P10) = \alpha_L + \beta_L \cdot |\mu|$$

$$(4) \qquad (P90 - \mu) = \alpha_R + \beta_R \cdot |\mu|,$$

within quarter and at the respondent level. The equations are separately estimated for intervals around positive return forecasts ($\mu>0$) and for intervals around negative forecasts ($\mu<0$). Positive β_L and β_R coefficients suggest that the 80% confidence intervals expand, in both directions, as the return forecasts get more extreme (hypothesis 2a). If forecast hedging, moreover, intensifies with the extremity of the forecasts, then β_L should exceed β_R for the intervals around positive μ, while β_R should exceed β_L for the intervals around negative μ (hypothesis 2b). Table V summarizes the results of testing the two hypotheses, by presenting the mean estimated β_L and β_R coefficients in quarterly-level and respondent-level regressions (Web supplement D presents the full estimation results). Since the long-term forecasts were negative in only 56 cases, the LT estimations are only run on the intervals around positive μ. The results of all the comparisons, however, uniformly support the two hypotheses. The β_L and β_R coefficients are positive in all twelve cells of the table. In the estimations building on intervals around positive μ, the mean β_L is at least 1/4 larger than the mean β_R and equality is rejected at $p<0.01$ in all four comparisons. The directions reverse in the estimations building on the (ST) intervals around negative $\mu<0$, where the mean β_R is more than 2/3 larger than the mean β_L and equality is again rejected at $p<0.01$. Hypotheses 2a and 2b are therefore strongly supported. The responsiveness of the intervals to the extremity of the forecasts, however, is smaller at the respondent level compared to the quarterly level; e.g., mean β_L 0.61 in the ST $\mu>0$ quarterly estimations, compared to about half smaller 0.31 in the parallel respondent-level estimations.

Table V: The increase in forecast hedging with forecast extremity

	ST forecasts		LT forecasts	
	$\boldsymbol{\beta_L}$	$\boldsymbol{\beta_R}$	$\boldsymbol{\beta_L}$	$\boldsymbol{\beta_R}$
Quarterly level				
μ>0	**0.61*** (44,0)	0.34*** (43,0)	**0.60*** (44,0)	0.47*** (42,0)
μ<0	0.49*** (10,1)	**0.82*** (32,0)	-	-
Respondent level				
μ>0	**0.31*** (26%,4%)	0.14*** (19%,6%)	**0.28*** (34%,4%)	0.12*** (20%,9%)
μ<0	0.08 (14%,10%)	**0.77*** (57%,0%)	-	-

Notes: The table presents the mean β_L and β_R coefficients in estimating equations (3) and (4) at the quarterly level (upper panel) and the respondent level (lower panel). The OLS estimations are separately run on the intervals around positive and negative expected return forecasts. The errors are corrected for heteroskedasticity at the quarterly level regressions and corrected using the Newey and West procedure at the respondent level time series regressions. Quarters or respondents with less than five observations are ignored. The samples for the quarterly level ST and LT μ>0 regressions consist of all 50 quarters. The sample for the quarterly level ST μ<0 regressions consist of 45 quarters. The samples for the respondent level regressions consist of 844 (ST μ>0), 942 (LT μ>0), and 21 (ST μ<0) respondents. The table presents the mean estimated coefficients for the relevant quarters (upper panel) or the relevant respondents (lower panel). In the upper panel, the smaller font parentheses present the number of quarters with statistically significant ($p \leq 0.05$) positive and negative coefficients; e.g., in the ST μ>0 regressions β_L is positive and significant in 44 quarters. In the lower panel, the smaller font parentheses present the proportion of respondents with significant positive and negative coefficients; e.g., in the ST μ>0 regressions the β_L is positive (negative) and significant for 26% (4%) of the respondents. In both panels, a sign test is used to test the hypotheses $\beta_L=0$ or $\beta_R=0$, with the three asterisks representing rejection at $p \leq 0.01$. The larger mean coefficient, β_L or β_R, in the paired estimations of equations (3) and (4) on the same sample, is bolded for emphasis. A sign test is used to test the hypothesis $\beta_L - \beta_R = 0$, with the dark shading representing rejection at $p \leq 0.01$.

3.4: The increase in perceived volatility with forecast extremity, beyond forecast hedging

To summarize the responsiveness of σ to |μ|, we estimate the equation

$$(5) \quad \sigma = \alpha + \beta \cdot |\mu| + \gamma \cdot 1_{\mu \leq 0},$$

where $1_{\mu \leq 0}$ is an indicator for non-positive return forecasts which is included to account for the increased length of these intervals.[12] The top row of Table VI presents the mean estimated β in quarterly-level and respondent-level regressions. In the quarterly-level estimations, the β estimates are always positive, and statistically significant at $p \leq 0.05$ in 45 (ST) and 48 (LT) of the 50 quarters. The mean β is 0.35 in the estimations building on the

[12] The linear specification appears to fit the data best. Log transformations, adding squared μ, or the interaction $|\mu| * 1_{\mu \leq 0}$ do not improve the fit. The mean R^2 in the four estimations are 0.16, 0.30, 0.31, 0.28.

short-term forecasts, and 0.41 in the estimations building on the long-term forecasts. Given the mean σ values of 5.5 in ST and 3.8 in LT, the slopes appear steep. Indeed, the mean σ of the ST intervals around |μ|<5% is 3.8%, compared to more than twice larger mean σ of 8.5% for the ST intervals around |μ|>10%. The differences are larger for the LT intervals, where the respective mean σ values are 2.7% and 8.4%. Moving to the respondent-level regressions, the mean β estimates are still positive for more than 68% of the respondents, but the mean coefficients drop by more than 40% compared to the quarterly-level results. In both levels of analysis, however, replacing |μ| with the signed μ or defining forecast extremity with respect to historical returns decreases the fit levels significantly (see Web supplement D for details).[13]

Table VI: The increase in σ with forecast extremity

	Quarterly level		Respondent level	
	ST forecasts	**LT forecasts**	**ST forecasts**	**LT forecasts**
Model (5) β	0.35*** (45,0)	0.41*** (48,0)	0.19*** (25%,4%)	0.15*** (29%,6%)
Model (6) β	0.24*** (46,0)	0.28*** (45,0)	0.13*** (25%,5%)	0.11*** (26%,8%)

Notes: The table presents the mean β coefficients in estimating equations (5) and (6) at the quarterly level and the respondent level. The method of the regressions is similar to the method of Table V. The left panel presents the mean estimated coefficients for the 50 quarters. The number of observations per quarter ranges between 174 and 417, and the mean is about 283. The right panel presents the mean estimated coefficients for the 958 (ST) and 952 (LT) respondents with at least five observations. The number of observations per participant ranges between 5 and 44, with an average close to 9.9 in ST and in LT. The smaller parentheses present the number of quarters (in the left panel) or the proportions of respondents (in the right panel) with statistically significant (at p≤0.05) positive and negative coefficients. A sign test is used to test the hypothesis that the quarterly or individually estimated parameters are centered at zero, with the three asterisks denoting rejection at p≤0.01. An extended version of the table with the full regression results is provided in Web supplement D.

Given the asymmetric increase in the length of the intervals with forecast extremity, it is interesting to test to what extent the increase in σ with |μ| follows from the strengthening of forecast hedging. For this test, we symmetrize the forecast intervals around μ, decreasing the size of the left margin to the size of the right margin when (μ-P10)>(P90-μ) and vice

[13] When the μ and σ correlations are calculated for the observations taking value between 0 and the median μ for the respective quarter (taken as an estimate of the prevailing consensus) the correlations are mostly positive. The mean correlations are 0.07 in ST (37 positive; 13 negative) and 0.10 in LT (42 positive; 8 negative).

versa, decreasing the size of the right margin to the size of the left margin when (μ-P10)<(P90-μ). Using $\sigma_{\text{Skewness removed}}$ for the σ estimates derived from the redefined intervals, we estimate the equation

$$(6) \qquad \sigma_{\text{Skewness removed}} = \alpha + \beta \cdot |\mu| + \gamma \cdot 1_{\mu \leq 0},$$

at the quarterly and respondent levels. The estimated β coefficients are presented at the bottom row of Table VI. The coefficients are still positive in all 50 quarters and for more than 65% of the respondents, and the mean estimates fall by less than 1/3 compared to model (5). The intensification of forecast hedging with forecast extremity therefore loosely accounts for less than 1/3 of the |μ| and σ correlations, and the correlations are still robustly significant when the effect of increased skewness is removed. Similar results emerge when the symmetrization alternatively builds on the right (left) margin in cases of positive (negative) forecasts.[14]

3.5: Reversal of the correlations, across the 50 quarters

In both tables V and VI, the slopes of the respondent-level (time series) regressions are substantially smaller than the slopes of the quarterly-level regressions. The results in Amromin and Sharpe (2014) suggest that the decrease in slopes may partially follow from the contrasting effects of market sentiments on forecast optimism and perceived volatility. If strong sentiments positively affect the return forecasts while negatively affecting the volatility estimates, then it is even surprising to find a positive link between forecast extremity and perceived volatility, especially in LT, where 98.8% of the forecasts are positive.[15]

[14] As the beta distribution reduces to a uniform distribution on [L,H] when the interval is symmetric (*a*=*b*=1), $\sigma_{\text{Skewness removed}}$=1.25·(2·min(P90-μ,μ-P10))·sqrt(1/12). When the right [left] margin of the interval is alternatively used for the symmetrization, $\sigma_{\text{Alternative symmetrization}}$ = 1.25·(2·(P90-μ))·sqrt(1/12) when μ>0 and $\sigma_{\text{Alternative symmetrization}}$ = 1.25·(2·(μ-P10))·sqrt(1/12) when μ<0. The mean β coefficients for the alternative symmetrization are 0.26; 0.35; 0.12; 0.09 (following the order from left to right in the table). The less than 1/3 weight of forecast hedging also reflects in the 2·min(P90-μ,μ-P10)/(P90-P10) ratios that take values (0.72, 0.75, 0.72, 0.75), respectively.

[15] The weaker results at the respondent level may also relate to the small (5-44) number of observations in the respondent-level regressions. But note that the widest ST (LT) interval submitted by a given respondent is at least twice larger than the narrowest ST (LT) interval submitted by the same respondent in 66% (60%) of the cases, so the σ estimates still exhibit considerable variability at the respondent level.

The top rows of Table VII indeed show that when the forecasts and the perceived volatility estimates are averaged at the quarterly level, the perceived volatility appears to decrease with forecast optimism. The correlation between the average signed forecasts ($\bar{\mu}$) and the average volatility estimates ($\bar{\sigma}$) is negative -0.20 (p=0.17) based on the short-term forecasts and stronger -0.38 (p<0.01) based on the long-term forecasts. Both correlations strengthen when optimism is measured in terms of the proportion of forecasts exceeding given thresholds. The second row of the table presents the correlation between the proportion of forecasts exceeding 5% ($1_{\mu\geq5}$) and $\bar{\sigma}$. The ST correlation becomes -0.35 (p=0.02) and the LT correlation strengthens to -0.44 (p<0.01). The $\bar{\sigma}$ for the quarter with the most pessimistic ST forecasts (2009:Q1, where $\bar{\mu}$=1.7) is 7.8, compared to 4.7 for the quarter with the most optimistic forecasts (2007:Q2, where $\bar{\mu}$=7.6). Across the 50 quarters, we therefore find evidence fitting the literature on the misperception of the risk-return tradeoff (see Kaustia et al., 2009; Kempf et al., 2014 recent evidence). Similarly to Amromin and Sharpe (2014), the bottom rows of Table VII show that variables related to macro sentiments exhibit opposite effects on forecast optimism and perceived volatility across the 50 quarters. The results are strong for ST but weaker for LT. Since this line of results is secondary to the main interest of the current paper, we leave the more detailed discussion to Web supplement D.

Table VII: Reversal of correlations, across the 50 quarters

	ST forecasts		**LT forecasts**	
	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\mu}$	$\bar{\sigma}$
$\bar{\mu}$	-	-0.20	-	-0.38***
$1_{\mu\geq5}$	-	-0.35**	-	-0.44***
VIX	-0.40***	0.72***	0.25	0.34**
SENT	0.33**	-0.63***	-0.14	-0.44***
Recent SP500	0.52***	-0.21	-0.28*	0.09

Notes: $\bar{\mu}$ represents the quarterly average return forecasts, $\bar{\sigma}$ is the quarterly average perceived volatility, and $1_{\mu\geq5}$ is the proportion of forecasts exceeding 5% within each quarter. The first two rows of the table present the weighted correlations between $\bar{\mu}$ or $\overline{1_{\mu\geq5}}$ and $\bar{\sigma}$ across the 50 quarters of the survey, where the weighting is based on the number of quarterly observations. In the bottom rows of the table, *VIX* represents the closing level of the VIX index at the survey date, *SENT* is the Michigan Consumer Sentiment Index for the survey date, and *Recent SP500* is the SP500 return in the three months preceding the survey date. The table presents the $\bar{\mu}$ and $\bar{\sigma}$ weighted correlations with VIX, SENT and Recent SP500 across the 50 quarters, where the weighting is based on the number of quarterly observations. The ST (LT) columns present the correlations based on the short-term (long-term) forecasts. The asterisks summarize the results of permutation tests on the correlations, with one/two/three asterisks representing significance at p≤0.01/0.05/0.1.

4. The implications of the decrease in confidence with forecast extremity, assuming CPT preferences

This section explores the implications of the decrease in confidence with forecast extremity, building on the panel of SP500 forecasts. The analysis is confined to the 12,453 forecasts around positive $\mu>0$ (henceforth addressed as the "*basic panel*").[16] While an increase in the expected return should positively affect the willingness to invest in the index, the increase in perceived volatility that comes with more extreme forecasts, may have the opposite effect of decreasing the appeal of investment (e.g., Nosic and Weber, 2010; Merkle and Weber, 2014; Ameriks et al., 2020; Calvo-Pardo et al., 2021). Assuming cumulative prospect theory preferences, we use the certainty equivalents (CEs) of the beta distributions representing the forecasts to proxy the willingness to invest in the index. The analysis is separated into two main steps. First, we count how often, within each quarter or at the respondent level, the decrease in confidence with forecast extremity cancels off the response to more optimistic expectations, so that $\mu'>\mu$ but $CE'<CE$. We then use permutation tests to more generally examine if the decrease in confidence significantly weakens the response to optimistic expectations, across the quarterly samples and at the respondent level. The permutations are then additionally utilized to show that the decrease in confidence partially mitigates the underestimation of volatility and increases the willingness to protect against weak performance of the SP500. The use of exogenously fixed preferences for the analysis has the advantage of expunging all irrelevant confounds. In the field and in experimental setups, investment decisions depend on individual preferences, personal characteristics and experimental methodologies (e.g., Dorn and Huberman, 2005; Fellner and Maciejovsky, 2007; Charness et al., 2012). By using the panel forecasts of finance professionals as the basis for the analysis, but fixing the preferences using two diverse versions of CPT, the analysis here explores the implications of the decrease in confidence with forecast optimism in a synthetic, yet tightly controlled setting that overcomes the hurdles of running such tests using empirical or experimental

[16] The number of observations per quarter is 162-373 with mean 249. The number of respondents with at least five forecasts is 846, with 5-39 observations per respondent (mean 9.4).

data.[17] Section 4.1 explains how we derive the certainty equivalents from the forecasts. Sections 4.2-4.5 present the results.

4.1: Deriving the CEs of the return distributions, assuming CPT

Cumulative prospect theory, developed by Tversky and Kahneman (1992), is commonly considered the best descriptive theory for individual choice under conditions of risk or uncertainty (Wakker, 2010). In finance, the theory has been utilized to resolve paradoxes regarding the behavior of investors and shed light on anomalies in the financial markets (Barberis and Thaler, 2003; Barberis, 2018). Currently, we build on the estimates reported in Tversky and Kahneman (1992) and on the more recent L'Haridon and Vieider (2019) estimates for a sample of almost 3000 participants from 30 countries. The Tversky and Kahneman (1992; henceforth addressed as *TK92*) CPT model is the canonical version that has been used to resolve the equity premium puzzle and other anomalies in the financial markets. The L'Haridon and Vieider (2019; henceforth addressed as *L'HV19*) is selected from the more recent CPT estimations that point at lower levels of risk and loss aversions compared to the canonical estimates (see Brown et al., 2021 meta-analysis).[18] The exact structure of CPT and these the two specific versions is outlined in Web supplement E.

To proxy the CFOs' willingness to invest in SP500, we construct discrete approximations of the four-parameters beta distributions and apply CPT to derive the certainty equivalents of the return distributions, assuming TK92 or L'HV19. The certainty equivalent (CE) of a random return is defined as the certain return level that makes the decision maker indifferent between the certain and the random alternatives. Using V_{CPT} for a given CPT valuation model, the CE of the random return R, CE(R), is formally defined through the equation $V_{CPT}(CE(R))=V_{CPT}(R)$. The certainty equivalents of given random returns are useful proxies for the relative willingness to invest in the respective assets. In a binary choice between R1 and R2, the investor should, ignoring noise, prefer R1 to R2 if and only

[17] See relatedly the empirical literature discussion of the attenuation puzzle (Ameriks et al., 2020).

[18] Another advantage of these two CPT models for the current use is a symmetry of the gain and loss utility functions, that makes the CE proportional to the investment budget. The insights provided by CPT models assuming smaller loss aversion compared to the canonical estimates is illustrated in Barberis et al. (2021).

if the certainty equivalent of R1 is larger. We accordingly use the CPT approximations, to test how often CE'<CE although μ'>μ and run permutation tests of the hypothesis that a decrease in confidence delays the response to optimistic expectations.

The discrete approximations build on the 4% quantiles of the Beta(a,b,L,H) distributions, using the medians of the intervals between adjacent 4% quantiles to represent the discrete return levels. The result of the discretization is a 25-outcomes probability distribution, $\{(r_k, 0.04)\}_{k=1 \text{ to } 25}$, where r_k is the (2+(k-1)·4)% quantile of the beta distribution, and the 0.04 represents the probability of the r_k return. The CPT model is applied to the discrete probability distribution to derive its CPT value, assuming the TK92 or the L'HV19 models, and the certainty equivalents are derived from these values. Table IIX presents descriptive statistics for the forecasts in the basic panel and their certainty equivalents under the two alternative CPT parametrizations. The certainty equivalents assuming the L'HV19 parameters exceed the CEs implied by TK92 in 89.4% of the cases, but the correlation between the two CEs is close to perfect exceeding 0.98.[19]

Table IIX: The CEs of the basic panel observations

	Mean	STD	Proportion < 0	95% range
μ	6.1	3.8	-	[2, 12]
σ	5.1	4.3	-	[0.7, 14.3]
CE assuming TK92	3.7	3.9	18.3%	[-1.4, 9.9]
CE assuming L'HV19	4.4	4.4	13.4%	[-1.7, 11.1]

Notes: The table presents descriptive statistics for the N=12,453 observations in the basic panel and their certainty equivalents (CEs), derived as explained in the text. The percentage sign is suppressed when presenting the forecast-related variables.

4.2: Can the decrease in confidence offset the response to optimistic expectations?

Since cumulative prospect theory obeys first-order stochastic dominance, a shift to the right, by positive constant K, of the expected return forecast μ and the confidence limits

[19] A detailed example for the derivations is provided in Web supplement E. The L'HV19 CEs fall below the TK92 CEs when the expected return forecasts are low and/or the intervals are relatively long. The mean μ in the cases where the L'HV19 CEs are smaller is 3.9 compared to 6.4 in the cases where the TK92 CEs are smaller. The mean σ's are 10.5 and 4.5, respectively.

P10 and P90, must increase the certainty equivalent.[20] In the basic panel, however, an increase in μ typically comes with an increase in $(\mu - P10)$, a smaller increase in $(P90 - \mu)$, and an overall increase in the length of the intervals and the estimated σ. The increase in perceived volatility may counterbalance, and even fully offset, the response to more optimistic expectations, so that CE'<CE although μ'>μ.

To check if the change in perceived volatility is steep enough to offset the response to more optimistic expectations within the 50 quarters of the panel, we construct samples of all the same-quarter observations (i,t) and (j,t) with $\mu_{i,t} > \mu_{j,t}$ and check how often $CE_{i,t} < CE_{j,t}$ so that the willingness to invest decreases in spite of the more optimistic expectations. The results reveal that, on average, the certainty equivalents decrease in about 20% of the comparisons. The CEs derived assuming the TK92 preferences decrease, on average, in 22.1% of the cases. The CEs derived assuming the L'HV19 preferences decrease, on average, in 18.4% of the cases.[21]

The analysis is also be applied at the respondent level, constructing all the same-respondent observations (i,t1) and (i,t2) with $\mu_{i,t1} > \mu_{i,t2}$ and checking how often $CE_{i,t1} < CE_{i,t2}$ so that the willingness to invest decreases in spite of the stronger optimism. Again, we restrict the individual-level tests to respondents with at least five observations (N=844 respondents; 7,932 basic panel observations). The results here are weaker compared to the quarterly-level analysis, matching the weaker increase in σ with μ at the individual level. The CEs derived assuming the TK92 preferences decrease, on average, in 14.3% of the cases, and the CEs derived assuming L'HV19 decrease, on average, in 12.2% of the cases.[22]

[20] Dominance is preserved in the approximations. The shift by K does not affect the (μ-P10) and (P90-μ) distances, so the *a*,*b* are not affected, and since the q-th quantile of Beta(*a*,*b*,L,H) is L+(H-L)*(q-th quantile of Beta(*a*,*b*)), the return levels in the discrete approximation also increase by K.

[21] The proportion of pairs with $\mathrm{CE}_{\mathrm{i,t}} < \mathrm{CE}_{\mathrm{j,t}}$ is calculated for each quarter and we report the mean proportion across the 50 quarters. The average number of comparisons per quarter is about 28,492.

[22] The proportion of pairs with $\mathrm{CE}_{\mathrm{i,t}} < \mathrm{CE}_{\mathrm{j,t}}$ is calculated for each respondent with at least five positive expected return forecasts (N=844), and we report the mean proportion across the sample. The average number of comparisons per respondent is about 45.

4.3: Permutation tests

In the basic panel, the response to more optimistic forecasts (within quarters or at the respondent level) may weaken because of the parallel changes in the forecast intervals. The current section reports the results of permutation tests that dissolve the |μ| and σ correlations by rematching the margins of the forecast intervals with the excepted returns within-quarter or at the respondent level. The observations in the permuted panels have the form (μ,[μ-(μ'-P10'),μ+(P90'-μ')]), where the apostrophe represents the observation that was randomly matched with the observation around μ. We derive the certainty equivalents of the permuted panels' observations and compare the responsiveness of the certainty equivalents to the expected return in the basic panel and the permuted panels. The permuted panels are henceforth addressed as *P-panels* to distinguish from the *basic panel*.

For the quarterly-level permutation tests, the equation $CE = \alpha + \beta \cdot \mu$ is first estimated separately for each of the 50 quarters. The average slope across the 50 quarters, denoted $\overline{\beta_{real}}$, is used to summarize the responsiveness of the certainty equivalents to the expected return in the basic panel. For the permutations, the program generates observations of the type $(\mu_{i,t};\ [P10'_{i,t}, P90'_{i,t}])$, where $P10'_{i,t} = \mu_{i,t} - (\mu_{j,t} - P10_{j,t})$, $P90'_{i,t} = \mu_{i,t} + (P90_{j,t} - \mu_{j,t})$, and the index j represents the quarter-t respondent that is randomly matched with respondent i. The beta distributions and certainty equivalents (*CE'*) are derived for the P-panel observations following the same method as applied for the observations in the basic panel.[23] The equation $CE' = \alpha + \beta \cdot \mu$ is then estimated separately for each of the 50 quarters, and the mean slope across the 50 quarters, denoted $\overline{\beta_P}$, is used to summarize the responsiveness of the certainty equivalents to the expected return in the P-panel. The procedure is repeated 1000 times, and the proportion of P-panels with $|\overline{\beta_P}| \leq |\overline{\beta_{real}}|$ is used to test if the decrease in confidence with forecast extremity significantly delays the responsiveness to optimistic expectations.

[23] In particular, $\mathrm{L}'_{\mathrm{i,t}} = \mu_{\mathrm{i,t}} - 1.25 \cdot (\mu_{\mathrm{j,t}} - \mathrm{P10}_{\mathrm{j,t}})$ and $\mathrm{H}'_{\mathrm{i,t}} = \mu_{\mathrm{i,t}} + 1.25 \cdot (\mathrm{H90}_{\mathrm{j,t}} - \mu_{\mathrm{j,t}})$. Note however that since $\mu_{i,t} - P10'_{i,t} = \mu_{j,t} - P10_{j,t}$ and $P90'_{i,t} - \mu_{i,t} = P90_{j,t} - \mu_{j,t}$, $a'_{i,t} = a_{j,t}$ and $b'_{i,t} = b_{j,t}$, so that the procedure basically transfers the shape parameters of j to i.

The left columns of Table IX summarize the results of the test. When the certainty equivalents are derived assuming the TK92 preferences, $\overline{\beta_{real}}$ is 0.77. In the basic panel, the TK92 certainty equivalents thus increase on average by 7.7% when the expected return forecast increases by 10%. The respective mean $\overline{\beta_P}$ in 1000 permutations however is 0.97, suggesting that the certainty equivalents increase on average by 9.7% for 10% increase in μ. The responsiveness of the TK92 certainty equivalents to the expected return thus increases by more than 1/4, from 0.77 to 0.97, when the μ and σ correlations are dissolved. Moreover, the 1000 P-panels $\overline{\beta_P}$ are almost always larger than $\overline{\beta_{real}}$, so the hypothesis that the decrease in confidence with forecast extremity weakens the response to optimistic expectations is supported at $p<0.01$. When the certainty equivalents are derived assuming the L'HV19 model, the gap between $\overline{\beta_{real}}$ and the mean $\overline{\beta_P}$ in the 1000 P-panels narrows to 0.15 (see the respective column of Table IX). However, as in the tests based on the TK92 CEs, $\overline{\beta_P}$ almost always exceeds the $\overline{\beta_{real}}$, so the hypothesis is supported again at $p<0.01$.

Table IX: Results of permutation tests

	Quarterly level		**Respondent level**	
	TK92	**L'HV19**	**TK92**	**L'HV19**
Mean slope in the basic panel	0.77	0.90	0.86	0.98
Mean slope in 1000 P-panels	0.97	1.05	0.96	1.06
Significance (% $\lvert\overline{\beta_P}\rvert \le \lvert\overline{\beta_{real}}\rvert$)	$p<0.01$	$p<0.01$	$p<0.01$	$p<0.01$

Notes: The equation CE=α+β·μ is estimated for each quarter (N=50) and each respondent with at least five basic panel observations (N=844), assuming the TK92 and the L'HV19 certainty equivalents. The top row of the table presents the mean $\overline{\beta_{real}}$ slopes in the 50 quarterly-level regressions (left columns) and in the 844 respondent-level regressions (right columns). The intermediate row presents the respective mean slopes in 1000 P-panels ($\overline{\beta_P}$), constructed using the method discussed in the text. The bottom row summarizes the results of the tests, reporting the proportion of P-panels where $\lvert\overline{\beta_P}\rvert \le \lvert\overline{\beta_{real}}\rvert$.

The right columns of Table IX show that similar conclusions emerge when the permutation tests are run at the respondent level. For the respondent-level tests, the $CE = \alpha + \beta \cdot \mu$ equation is estimated for each respondent with at least five basic panel observations and $\overline{\beta_{real}}$ is calculated by averaging the β slopes across the N=844 respondents. The P-panel observations now have the structure ($\mu_{i,t}$; $[P10'_{i,t}, P90'_{i,t}]$), where $P10'_{i,t} = \mu_{i,t} - (\mu_{i,t'} - P10_{i,t'})$, $P90'_{i,t} = \mu_{i,t} + (P90_{i,t'} - \mu_{i,t'})$, and the index t' represents the quarter that is randomly matched with quarter t in a given permutation for respondent i. The P-panels $\overline{\beta_P}$ are again obtained by deriving the certainty equivalents, re-running the regressions, and averaging the slopes across the sample of 844 respondents. The results reveal that $\overline{\beta_P}$ is 0.10 (assuming TK92) or 0.08 (assuming L'HV19) larger than $\overline{\beta_{real}}$. The hypothesis that the decrease in confidence weakens the responsiveness of CPT investors to optimistic expectations is again supported at $p<0.01$.

4.4: The decrease in confidence alleviates miscalibration

Ben-David, Graham and Harvey (2013; henceforth BGH13) analyze the SP500 return forecasts submitted by the Duke panel CFOs up to 2011:Q1, showing that the forecasts are severely miscalibrated and miscalibration links with more aggressive corporate policies. The volatility estimates derived from the intervals analyzed in BGH13 are more than 2/3 smaller than the VIX for the survey date, suggesting that the CFOs vastly underestimate the index volatility. The realized SP500 returns fall within the 80% confidence intervals in only 36% of the cases. The current paper does not explore issues related to miscalibration, except for this section, where we show that the decrease in confidence with forecast extremity partially lightens the underestimation of volatility and improves calibration.[24]

Table X presents the results of splitting the basic panel sample depending on the extremity of the positive return forecasts. Again, we control for the panel structure of the data by reporting the results of quarterly-level and respondent-level analyses. In both levels of analysis, the increase in σ with forecast optimism reduces the underestimation of volatility. The mean quarterly-level σ /VIX ratio is 0.23 for the $0<\mu\leq4$ intervals, 0.30 for the intervals around $4<\mu\leq8$, and 0.42 for the $\mu>10$ intervals, and the individual-level analysis shows similar results. The hit rates of the intervals, defined as the proportion of cases where the realized return falls within the forecast interval, also increase as the forecasts turn more optimistic. The mean quarterly-level hit rate of the intervals around $0<\mu\leq4$ is about 17%, compared to 27% for the $4<\mu\leq8$ intervals, and 48% for the intervals around $\mu>8$, and again

[24] The overconfidence rates in our sample are similar to the rates reported in BGH13. Our mean σ is about 5.5 compared to a respective mean VIX of 19.0, while BGH13 report mean σ 5.5 and mean VIX 22. The hit rate in our sample is about 27%, compared to the 36% rate in BGH13. A closer look reveals hit rates close to 0% for 2007:Q4, 2008:Q1, 2012:Q4 and 2016:Q1 (See Web supplement D for more details).

the individual-level results are similar. Technically, the increase in hit rates with the extremity of the forecasts can follow from more accurate forecasts or from the increase in the length of the intervals (Sonsino et al., 2021). To disentangle the effects, we use the P-panels that randomly match the margins of the intervals with the return forecasts across the quarters or at the respondent level. Table X shows that the hit rates of the intervals around $\mu>8$ significantly decrease (by 12.5% at the quarterly level; 8.7% at the respondent level) in the P-panels compared to the basic panel. The decrease in confidence with optimism thus boosts the calibration rates of the optimistic CFOs, controlling for the accuracy of their forecasts. Interestingly, the P-panels hit rates consistently increase from left to right in both levels of analysis. Such increase can only follow from an increase in the accuracy of the expected return forecasts. The "|Error|" and "Relatively small |Error|" rows of the table indeed show that the absolute prediction errors decrease and the proportion of relatively small prediction errors increases with the extremity of the forecasts (see the table notes for the variable definitions). While the improved calibration of the more optimistic CFOs may be specific to the sample, the permutation tests more interestingly suggest that the decrease in confidence with optimism alleviates miscalibration and improves the forecast quality of optimistic CFOs. Even the optimistic CFOs, however, underestimate the VIX volatility by more than 50%.

Table X: The decrease in miscalibration with optimism

	Full μ>0 sample	0<μ≤ 4	4<μ≤ 8	μ>8
Quarterly level				
σ	5.1	3.8	5.1	7.5
VIX	19.0	19.0	19.0	19.0
σ/VIX	0.30	0.23	0.30	0.42
Basic panel hit	27.4%	17.1%	27.0%	47.5%
P-panels hit	26.9%	22.4%	26.9%	35.0%
|Error|	12.8	13.7	12.5	11.8
Relatively small |Error|	33%	24%	33%	42%
Respondent level				
σ	5.2	4.3	5.0	7.5
VIX	19.7	21.0	18.8	20.8
σ/VIX	0.29	0.24	0.30	0.40
Basic panel hit	27.6%	16.7%	27.2%	49.0%
P-panels hit	27.5%	19.0%	28.0%	40.3%
|Error|	12.8	15.5	12.3	10.3
Relatively small |Error|	32%	24%	32%	46%

Notes: The basic panel sample is split depending on whether the expected return does not exceed 4% (the "0<μ≤4" column), falls between 4% and 8% ("4<μ≤ 8") or exceeds 8% ("μ>8"). The upper panel of the table presents the results of a quarterly-level analysis based on the 12,634 basic panel forecasts. The lower panel presents the results of a respondent-level analysis, which is restricted to the 844 respondents with at least five positive forecasts (7,932 observations). σ is the volatility estimate as defined previously. *VIX* is the closing level of the VIX index at the survey date. *Basic panel hit* is an indicator that takes value 1 when the realized SP500 return (in the year starting at the survey date) falls within the basic panel forecast interval. *P-panel hit* is an indicator that takes value 1 when the realized SP500 return falls within the P-panel forecast interval. |*Error*| is the absolute value of the difference between the expected return forecast (μ) and the realized SP500 return. *Relatively small* |*Error*| is an indicator that takes value 1 when |Error| is smaller than 0.5 the average length of the intervals submitted in the respective quarter (for the quarterly-level analysis) or smaller than 0.5 the average length of the intervals submitted by the respective respondent (for the respondent-level analysis). For the quarterly-level panel, the average values of the variables are calculated for each quarter and the table presents the means for the 50 quarters. For the respondent-level panel, the averages are calculated for each respondent with observations at the respective μ range, and the table presents the means for the relevant respondents. The vertically adjacent *Basic panel hit* and *P-Panels hit* are shaded when the permutation test rejects equality of the hit rates at p≤0.01. The sample sizes for each column are presented in a Web supplement E extended version of the table.

4.5: Implications – an example

This section presents an example illustrating that the decrease in confidence with optimism may boost the optimistic CFOs' willingness to protect against weak performance of the SP500 index. The levels of protection, however, are still about 2/3 smaller than implied by VIX-based estimates of the SP500 volatility.

The example refers to an investor that considers the choice between direct investment in the SP500 index and alternative investment in a one-year structured note that guarantees 97% of the investment capital and participates in 50% of the SP500 return beyond the -3% threshold.[25] Using the symbol SN for the return on the structured note, SN=-3%+0.5·Max(SP500+3%, 0%). If SP500 decreases by more than 3%, capital protection applies and the note pays 97% of the invested capital. The rate of capital protection increases when SP500 falls by less than 3%, and the return turns positive when the SP500 return exceeds 3%. Using V(SP500) for the CPT value of direct investment in SP500 and V(SN) for the CPT value of the structured note, we adopt the Fechner model with heteroskedastic noise (e.g., Blavatskyy and Pogrebna, 2010) to estimate the probability that the capital protecting note is preferred to direct investment in the index.[26] In particular, assume that the structured note is preferred to direct investment in the index when V(SN) + ε ≥ V(SP500), where ε ~ N (0, σ·(P90-P10)). The ε represents the zero-mean noise in the valuation process, with the parameter σ representing the baseline noisiness and the multiplication by (P90-P10) is the heteroskedastic adjustment. The probability that the structured note is preferred to direct investment in SP500 under this structure is

$$(7) \quad \text{Prob (SN preferred)} \equiv \Phi\left(\frac{V(SN) - V(SP500)}{\sigma \cdot [P90 - P10]}\right),$$

where Φ represents the standard normal cumulative distribution function. Based on Sonsino et al. (2021; Table 5 model (d)), we assume σ=0.2 and use equation (7) to derive the Prob(SN preferred) for each of the N=1,129 basic panel observations around an expected return forecast of 10%. The *Basic panel* rows of Table XI present the results of the quarterly-level and respondent-level analysis. Assuming the TK92 preferences, the mean quarterly-level probability that the structured deposit is preferred to direct investment in SP500 is 19.6%. Assuming the L'HV19 parameters, the respective mean probability is about 1/4 smaller, 14.7%. The results of the respondent-level analysis are comparable.

[25] Normalizing the SP500 price to 100, a duplicating portfolio (e.g., Stoimenov and Wilkens, 2005) for the note consists of a one-year bond with a face value of 97 and 0.5 fraction of a one-year SP500 option with a strike of 97. Assuming a risk-free rate of 3% and SP500 volatility of 20% (values close to the averages along the survey), the price of the bond is 94.2 and the price of the call option is about 11.0, so the fair value of the note is about 99.7.

[26] For diverse applications of the Fechner model with heteroskedastic noise see Buschena and Zilberman (2000), Wilcox (2008) and Alempaki et al. (2019).

Table XI: The probability that the structured note is preferred to SP500

	TK92	**L'HV19**
Quarterly level		
Basic panel	19.6% [6%,44%]	14.7% [1%,42%]
1000 P-panels	11.6% [0%,45%]	8.2% [0%,44%]
Extended intervals (σ=VIX)	56.3% [29%,77%]	53.0% [17%,85%]
Respondent level		
Basic panel	20.3% [0%,75%]	15.3% [0%,82%]
1000 P-panels	16.4% [0%,84%]	12.0% [0%,93%]
Extended intervals (σ=VIX)	57.9% [19%,95%]	55.4% [6%,100%]

Notes: The probability that a structured note with the payoff function -3%+0.5·MAX (SP500+3%, 0%) is preferred to direct investment in the SP500, in cases where the SP500 annual return forecast is 10%, is calculated using equation (7). The upper panel of the table presents the results of a quarterly-level analysis, building on the 1,129 observations around μ=10%. The average *Prob (SN is preferred)* is calculated for each quarter, and the table presents the mean across the 50 quarters. The lower panel presents the results of a respondent-level analysis, confined to the N=365 respondents with at least five positive forecasts, including at least one 10% forecast (751 μ=10% observations). The average *Prob (SN is preferred)* is calculated for each respondent, and the table presents the mean for the 365 respondents. The *Basic panel* rows present the mean *Prob (SN is preferred)* probabilities for the basic panel. The *1000 P-panels* rows present the mean probabilities for the 1000 P-panels. The *Extended intervals (σ=VIX)* rows present the mean probabilities for a modified version of the basic panel, where the P90-P10 intervals are extended to increase σ to the level of VIX at the survey date. The smaller brackets present the minimal and maximal values of *Prob (SN is preferred)* across the 50 (quarterly-level) or 365 (respondent-level) observations. In the *1000 P-panels* rows, the smaller brackets present the minimal and maximal values across the 1000 P-panels.

To quantitatively estimate the effect of the relatively low confidence in optimistic 10% forecasts on the willingness to invest in the structured note, we use the 1000 P-panels that rematch the margins of the intervals with the expected return forecasts. Table XI shows that the probability that the structured note is preferred to the SP500 is smaller in the P-panels compared to the basic panel, in all four comparisons. The quarterly-level TK92 calculations suggest that Prob(SN preferred) increases from 11.6% (in the P-panels) to 19.6% (in the basic panel), because of the relatively low confidence in optimistic SP500 forecasts of 10%. The probabilities assuming the L'HV19 preferences are 8.2% and 14.7%, respectively. In proportional terms, the quarterly-level permutations suggest that willingness to protect against weak performance of SP500 increases by more than 2/3 due to the decrease in confidence with optimism. The differences are smaller in the respondent-level analysis, but the Prob(SN preferred) is still about 25% higher in the basic panel compared to the P-panels.

Note, however, that the basic panel Prob(SN preferred) take low values, around 15% to 20%, in spite of the boosting effect of the low confidence. The *Extended intervals (σ=VIX)* rows of Table XI finally show that when the left and right margins of the forecast intervals are increased proportionally, to inflate σ to the level of VIX at the survey date, the mean allocations to structured note increase to levels exceeding 50%.[27] The willingness to invest in the structured note, in terms of the mean Prob(SN preferred) estimates, under VIX-based estimates of the volatility is roughly three times larger than the willingness to invest implied by the actual forecast intervals.

5. Empirical tests

This section turns to exploring the correlation between the extremity of realized returns and estimates of the concurrent volatility in historical return series. The analysis is run on the stocks that were members of the S&P 500 list at the end of 2021, but started trading before the beginning of 2000.[28] The list of 378 stocks is henceforth addressed as *the (stock) list*. The period between the start of January 2000 and the end of December 2021 is *the test window*. The realized return (*r*) on each stock in the list, along the test window, is measured at five levels: daily, monthly, quarterly, half-yearly and yearly. The contemporaneous volatility (σ) is estimated from the sub-period returns as explained in Table XII.

[27] Specifically, the lower (upper) bounds of the interval are changed to P10'=μ-(VIX/σ)·(μ-P10) (P90'=μ+(VIX/σ)·(P90-μ)). The shape parameters of the estimated beta distributions do not change in such proportional shifts, so σ'=VIX.

[28] Similar results emerge in other forms of analysis; e.g., using the full list of S&P 500 stocks at the end of 2021 and counting the stocks that started trading after 2000 from their first trading day.

Table XII: Levels of empirical analysis

Realized return (*r*)	**Measure of volatility (*σ*)**
Daily	The daily high minus low spread
Monthly	STD of daily returns within the month
Quarterly	STD of weekly returns within the quarter
Half-yearly	STD of the six monthly returns within the half-year
Yearly	STD of twelve monthly returns within the year

Notes: The table presents the five levels of analysis in the empirical tests of the correlations between realized returns (r) and estimates of the contemporaneous volatilities (σ), as explained in the text. STD abbreviates standard deviation. In the quarterly level analysis, the last week of the quarter is extended to include the remaining days.

In each level of analysis, we report four correlations:

-ρ1 is the correlation between the absolute realized returns and the respective volatility estimates, $\rho(|r|,\sigma)$

-ρ2 is the correlation between the signed realized returns and the respective volatility estimates, $\rho(r,\sigma)$

-ρ3 is the correlation between the realized returns and the respective volatility estimates, based on the observations with positive returns; i.e., $\rho(r,\sigma)$, based on the observations with $r>0$

-ρ4 is the correlation between the absolute realized return and the respective volatility estimate, based on the observations with negative returns; i.e., $\rho(|r|,\sigma)$, based on the observations with $r<0$

The 5x4 correlation matrices generated by crossing the five levels of analysis with the four types of correlations are presented in the two panels of Table XIII.

The upper panel summarizes the results of a cross-sectional analysis, based on the periodical r and σ observations for the 378 stocks. In the daily-level analysis, the correlations between the daily |r| (or the signed r) and σ are calculated for each of the 5536 days along the test window. The table presents the mean correlations, with the proportion of statistically significant positive and negative correlations in smaller font parentheses. For the monthly level analysis, the correlations are similarly derived for each of the 264 calendar months in the test period. Again, the table presents the mean cross-sectional correlation with the proportion of months with statistically significant positive or negative correlations in smaller font parentheses. The quarterly/half-yearly/yearly mean cross-sectional correlations are similarly derived. The asterisks summarize the results of a sign-test on the full sample of cross-sectional correlations with 3/2/1 asterisks representing significance at p≤0.01/0.05/0.1.

The lower panel of the table summarizes the results of a time-series analysis that is applied to each of the stocks in the list separately. In the daily-level analysis, the correlations between the daily r (or |r|) and σ are calculated for each stock in the list, based on the 5536 daily observations. The table presents the mean correlations, with the proportion of statistically significant positive and negative correlations in smaller font parentheses. The monthly/quarterly/half-yearly/yearly mean time-series correlations are similarly derived. The asterisks summarize the results of a sign test on the 378 correlations with 3/2/1 asterisks representing significance at p≤0.01/0.05/0.1.

Table XIII: The realized-return and contemporaneous volatility correlations

Cross-sectional correlations				
	ρ1	ρ2	ρ3	ρ4
Daily	0.16*** (70%, 0%)	0.01*** (23%, 19%)	0.18*** (54%, 0%)	0.18*** (55%, 0.1%)
Monthly	0.49*** (100%, 0%)	0.02 (40%, 33%)	0.47*** (94%, 0%)	0.53*** (97%, 0%)
Quarterly	0.42*** (99%, 0%)	0.03 (43%, 30%)	0.41*** (92%, 0%)	0.46*** (93%, 0%)
Half-yearly	0.33*** (100%, 0%)	0.04** (36%, 30%)	0.31*** (93%, 2%)	0.37*** (84%, 0%)
Yearly	0.34*** (96%, 0%)	0.04 (46%, 27%)	0.35*** (96%, 0%)	0.46*** (91%, 0%)
Time-series correlations				
	ρ1	ρ2	ρ3	ρ4
Daily	0.40*** (99%, 0%)	-0.02*** (5%, 37%)	0.39*** (98%, 0%)	0.42*** (98%, 0%)
Monthly	0.49*** (100%, 0%)	-0.12*** (3%, 53%)	0.45*** (99%, 0%)	0.58*** (99%, 0%)
Quarterly	0.41*** (90%, 0%)	-0.18*** (2%, 48%)	0.35*** (65%, 0%)	0.58*** (90%, 0%)
Half-yearly	0.22*** (34%, 0%)	-0.15*** (4%, 26%)	0.20*** (25%, 0.3%)	0.42*** (34%, 0%)
Yearly	0.17*** (17%, 0.5%)	-0.13*** (4%, 10%)	0.21*** (19%, 0.3%)	0.44*** (13%, 0%)

Notes: The table summarizes the results of testing the correlation between the extremity of realized returns and estimates of the concurrent volatility, based on empirical data. The analysis is run at five levels. The "**Daily**" analysis tests the correlations between the realized daily returns and the respective daily high-low spreads. "**Monthly**" tests the correlations between the realized monthly returns and the standard deviations of the daily returns within the respective months. "**Quarterly**" tests the correlations between the realized quarterly returns and the standard deviations of the weekly returns within the respective quarters. "**Half-yearly**" tests the correlations between the realized half-year returns and the standard deviations of the six monthly returns within the respective half-years. "**Yearly**" tests the correlations between the realized yearly returns and the standard deviations of the twelve monthly returns within the respective years. For each level of analysis, the table presents four correlations. **ρ1** is the correlation between the absolute realized returns and the respective volatility estimates. **ρ2** is the correlation between the signed realized returns and the respective volatility estimates. **ρ3** is the correlation between the realized returns and the respective volatility estimates, based on the subsamples of positive realized returns. **ρ4** is the correlation between the absolute realized returns and the respective volatility estimates, based on the subsamples of negative realized returns. The analysis is run on the 378 stocks that were members of the S&P 500 list at the end of 2021, but started trading before the beginning of 2000 (*the list*). The upper panel summarizes the results of a cross-sectional analysis, where the correlation is calculated for each period (day, month, quarter, half year, or year – depending on the level of analysis), between the beginning of 2000 until the end of 2021 (*the test window*). The daily analysis is based on the 5536 trading days along the test window. The monthly/quarterly/half-yearly/yearly analyses are based on 264/88/44/22 non-overlapping trading periods. The table presents the mean correlations, with the proportion of statistically significant ($p \leq 0.05$) positive (left) and negative (right) cross-sectional correlations in smaller font parentheses. The lower panel summarizes the results of a time-series analysis, applied to each of the 378 stocks in the list separately, based on the realized return series along the text window. The table presents the mean correlations, with the proportion of statistically significant ($p \leq 0.05$) positive (left) and negative (right) correlations in smaller font parentheses. In both types of analysis, correlations are counted only if based on at least five observations. The asterisks summarize the results of a sign test on the full set of correlations, with 3/2/1 asterisks representing significance at $p \leq 0.01/0.05/0.1$. Web supplement F presents an extended version of the table with more details on sample sizes.

More details on the analysis and the results are presented in Web supplement F. Currently, we briefly note that in all five levels of analysis and in both cross-sectional and time-series analyses, the absolute realized returns increase with the estimates of contemporaneous volatility (ρ1), and the correlations dissolve when the absolute return series are replaced by the signed return series (ρ2). Moreover, the increase in absolute returns with volatility separately shows for the positive sub-samples (ρ3) and the negative sub-samples (ρ4). We are not aware of preceding studies documenting a similar correlational pattern in empirical return series. Connecting the results to statistical properties of the cross-sectional stock returns or dynamic models of the stock returns is beyond the scope of the current explorative examination. We end this brief section noting that the behavioral correlations of the preceding section have empirical roots. Relatively extreme, positive or negative, realized returns show with higher contemporaneous volatility, at the cross-section and across time.

6. Discussion

Three experiments and a large panel of return forecasts revealed that forecast confidence decreases as the forecasts diverge from zero. The decreased confidence reflects in longer forecast intervals, for confidence levels taking values of 90% (experiment 1), 50% (experiment 2), and 80% (the CFOs panel). It also shows directly, when MBA students provide likelihood assessments to small intervals around their point predictions (experiment 3). The perceived volatility estimates derived from the CFOs' forecast intervals increase with the absolute expected returns at the quarterly level and also at the individual level, in spite of a negative correlation between the expected returns and the perceived volatilities across the 50 quarters.

The decrease in confidence with forecast extremity relates to behavioral theories that have been utilized to shed light on the judgment and decision-making of investors and households. De Bondt's (1993) forecast hedging theory argues that forecasters skew their return distributions consistently, exhibiting negative skewness in cases of positive forecasts and conversely showing positive skewness when the forecasts are negative. Forecast hedging currently showed in the CFOs' forecasts of the short-term and long-term SP500 annual return, and it also emerged in a computerized experiment where quartile forecasts for the FTSE annual return were elicited indirectly using the exchangeability method (experiment 2). Semantically, the term *forecast hedging* has a strategic connotation, suggesting that forecasters deliberately hedge their optimistic/pessimistic predictions. The results of experiment 2, where the forecasts were elicited indirectly in sequences of binary choice problems, suggest that contrarian hedging more basically characterizes the expectations formation of financially competent forecasters. In both data sets, experiment 2 and the panel, we moreover find that the contrarian hedging intensifies with the extremity of the forecasts. The skewness to the left increases as the forecasts turn more optimistic, and skewness to the right emerges for strongly pessimistic forecasts. The decrease in confidence with forecast extremity, however, is more general than forecast hedging, as the correlations between the absolute return forecasts and the volatility estimates decrease by only 1/3 and show clear significance when the confidence intervals are symmetrized around the expected return forecasts.

The decrease in confidence with forecast extremity also links to the literature on *proportional or relative thinking* (Thaler, 1980; Tversky and Kahneman, 1981; Azar, 2007; Bushong et al., 2021). Relative thinking models assert that decision-makers respond to relative differences, beyond absolute differences, in economic decisions. In Tversky and Kahneman's (1981) pioneering example, 2/3 of the subjects are willing to drive 20 minutes to save five dollars on a $15 calculator, while less than 1/3 chose to do so when the five-dollar saving applies to a $125 calculator. In the context of financial forecasting, a fixed length interval around a given point prediction, may appear relatively smaller as the point prediction gets more extreme. The likelihood that forecasters assign to return falling within the interval accordingly decreases with the absolute forecast. The intuition still applies when the forecasters independently generate confidence intervals for given returns. Tversky and Kahneman's (1974) *anchoring and adjustment theory* (AAT) argues that decision-makers generate confidence intervals for unknown quantities by adjusting from preliminary anchors. The adjustments are insufficient and the confidence intervals are too tight, leading to miscalibration. In financial forecasting, the natural anchor is a point prediction for the target return, and the insufficient adjustments show in confidence limits that tilt towards the point prediction (Harvey, 2007). Adding a flavor of proportional thinking to AAT, it is plausible however to assume that the adjustments increase with the size of the anchor, so the confidence intervals get longer as the point predictions diverge from zero. Anchoring and adjustment, augmented by proportional thinking, may technically increase the length of extreme interval forecasts.

Beyond the behavioral explanations, an exploratory analysis of the cross-sectional and time series correlations between realized returns and estimates of the contemporaneous realized volatility revealed robust correlations between the extremity of the realized returns, in the positive or negative directions, and the realized volatility estimates. As in the behavioral analysis, the correlations disappear when the absolute returns are replaced by the signed variables. Exploring these correlations in light of statistical or stochastic models of stock returns is beyond the scope of this paper. An instant intuition, for example, suggests that the correlations may relate to cross-sectional or time series variations in illiquidity

(Amihud, 2002). For the current use, we briefly note that beyond the behavioral explanations, the forecast extremity and perceived volatility correlations appear to exhibit empirical roots.

The decrease in confidence with forecast extremity also relates to the empirical literature on the trading styles of individual investors and the profitability of momentum strategies. Surveys and experiments show that private investors extrapolate trends when forecasting financial returns (e.g., De Bondt, 1993; Dominitz and Manski, 2011; Greenwood and Shleifer, 2014; Bao et al., 2021). Empirical studies, however, discover that private investors tend for contrarian trading rather than chasing trends (see Economou et al., 2022 comprehensive survey). In Griffin et al. (2003) investors exhibit contrarian response to the Nasdaq-100 daily returns, and Kaniel et al. (2008) report contrarian trading with respect to the NYSE monthly returns. Contrarian response to recent quarterly returns is reported for Finnish households (Grinblatt and Keloharju, 2000), German private investors (Baltzer et al., 2019), and Indian retail investors (Chhimwal and Bapat, 2021). The literature moreover argues that institutional investors benefit from momentum or feedback trading at the expense of private investors (Grinblat and Han, 2005; Barber et al., 2009; Baltzer et al., 2019; Economou et al., 2022). Connecting to the current results, the decrease in confidence with forecast extremity may deter private investors from chasing trends, explaining the prevalence of contrarian trading in empirical records. It may also explain the persistence of the momentum anomaly in spite of the interest of institutional investors (Israel and Moskowitz, 2013).

References

Abbas, A. E., Budescu, D. V., Yu, H. T., and Haggerty, R. (2008). A comparison of two probability encoding methods: Fixed probability vs. fixed variable values. *Decision Analysis*, *5*(4), 190-202.

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review*, *101*(2), 695-723.

Alempaki, D., Canic, E., Mullett, T. L., Skylark, W. J., Starmer, C., Stewart, N., and Tufano, F. (2019). Reexamining how utility and weighting functions get their shapes: A quasi-adversarial collaboration providing a new interpretation. *Management Science*, *65*(10), 4841-4862.

Ameriks, J., Kézdi, G., Lee, M., and Shapiro, M. D. (2020). Heterogeneity in expectations, risk tolerance, and household stock shares: The attenuation puzzle. *Journal of Business and Economic Statistics*, *38*(3), 633-646.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31-56.

Amromin, G., and Sharpe, S. A. (2014). From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science*, *60*(4), 845-866.

Azar, O. H. (2007). Relative thinking theory. *The Journal of Socio-Economics*, *36*(1), 1-14.

Baillon, A. (2008). Eliciting subjective probabilities through exchangeable events: An advantage and a limitation. *Decision Analysis*, *5*(2), 76-87.

Baltzer, M., Jank, S., and Smajlbegovic, E. (2019). Who trades on momentum?. *Journal of Financial Markets, 42*, 56-74.

Bao, T., Hommes, C., and Pei, J. (2021). Expectation formation in finance and macroeconomics: A review of new experimental evidence. *Journal of Behavioral and Experimental Finance, 32*, 100591.

Barber, B. M., Lee, Y. T., Liu, Y. J., and Odean, T. (2009). Just how much do individual investors lose by trading?. *The Review of Financial Studies*, *22*(2), 609-632.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 1, pp. 79-175). North-Holland.

Barberis, N., Jin, L. J., and Wang, B. (2021). Prospect theory and stock market anomalies. *The Journal of Finance*, *76*(5), 2639-2687.

Barberis, N., Thaler, R. (2003). "A survey of behavioral finance." In: Constantinides, G., Harris, M., Stulz, R. (Eds.), *Handbook of the Economics of Finance*, vol. 1. North Holland, Amsterdam.

Ben-David, I., Graham, J. R., and Harvey, C. R. (2013). Managerial miscalibration. *The Quarterly Journal of Economics*, *128*(4), 1547-1584.

Blavatskyy, P. R., and Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, *25*(6), 963-986.

Boutros, M., Ben-David, I., Graham, J. R., Harvey, C. R., and Payne, J. W. (2020*). The persistence of miscalibration* (No. w28010). National Bureau of Economic Research.

Brown, A. L., Imai, T., Vieider, F., and Camerer, C. (2021). Meta-analysis of empirical estimates of loss-aversion. *Available at SSRN 3772089*.

Buschena, D., and Zilberman, D. (2000). Generalized expected utility, heteroscedastic error, and path dependence in risky choice. *Journal of Risk and Uncertainty*, *20*(1), 67-88.

Bushong, B., Rabin, M., and Schwartzstein, J. (2021). A model of relative thinking. *The Review of Economic Studies*, *88*(1), 162-191.

Calvo-Pardo, H., Oliver, X., and Arrondel, L. (2021). Subjective Return Expectations, Perceptions, and Portfolio Choice. *Journal of Risk and Financial Management*, *15*(1), 6.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization*, *81*(1), 1-8.

Chhimwal, B., and Bapat, V. (2021). Comparative study of momentum and contrarian behavior of different investors: Evidence from the Indian Market. *Asia-Pacific Financial Markets*, *28*(1), 19-53.

De Bondt, W. P. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, *9*(3), 355-371.

Dominitz, J., and Manski, C. F. (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics*, *26*(3), 352-370.

Dorn, D., and Huberman, G. (2005). Talk and action: What individual investors say and what they do. *Review of Finance*, *9*(4), 437-481.

Du, N., and Budescu, D. V. (2007). Does past volatility affect investors' price forecasts and confidence judgements?. *International Journal of Forecasting*, *23*(3), 497-511.

Economou, F., Gavriilidis, K., Gebka, B., and Kallinterakis, V. (2022). Feedback trading: a review of theory and empirical evidence. Forthcoming, *Review of Behavioral Finance*.

Ederington, L. H., and Guan, W. (2006). Measuring historical volatility. *Journal of Applied Finance*, *16*(1).

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*, *27*(1), 30-41.

Eyting, M., and Schmidt, P. (2021). Belief elicitation with multiple point predictions. *European Economic Review*, *135*, 103700.

Fellner, G., and Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology*, *28*(3), 338-350.

Ganzach, Y. (2000). Judging risk and return of financial assets. *Organizational Behavior and Human Decision Processes*, *83*(2), 353-370.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1-2), 59-95.

Glaser, M., and Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1-36.

Graham, J., and Harvey, C. R. (2001). Expectations of equity risk premia, volatility and asymmetry from a corporate finance perspective. Mimeo.

Granger, C. W. J., and Ding, Z. (1995). Some properties of absolute return: An alternative measure of risk. *Annales d'Economie et de Statistique*, 67-91.

Greenwood, R., and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, *27*(3), 714-746.

Griffin, J. M., Harris, J. H., and Topaloglu, S. (2003). The dynamics of institutional and individual trading. *The Journal of Finance*, *58*(6), 2285-2320.

Grinblatt, M., and Han, B. (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*, *78*(2), 311-339.

Grinblatt, M., and Keloharju, M. (2000). The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of Financial Economics*, *55*(1), 43-67.

Groeneveld, R. A., and Meeden, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *33*(4), 391-399.

Grosshans, D., and Zeisberger, S. (2018). All's well that ends well? On the importance of how returns are achieved. *Journal of Banking and Finance*, *87*, 397-410.

Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking and Reasoning*, *13*(1), 5-24.

Hossain, T., and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*(3), 984-1001.

Israel, R., Moskowitz, T., (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108, 275–301.

Jiao, P. (2020). Payoff-based belief distortion. *The Economic Journal*, *130*(629), 1416-1444.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions, volume 2* (Vol. 289). John Wiley and sons.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, *57*(2), 226-246.

Kaniel, R., Saar, G., and Titman, S. (2008). Individual investor trading and stock returns. *The Journal of Finance*, *63*(1), 273-310.

Kaustia, M., Laukkanen, H., and Puttonen, V. (2009). Should good stocks have high prices or high returns?. *Financial Analysts Journal*, *65*(3), 55-62.

Kempf, A., Merkle, C., and Niessen-Ruenzi, A. (2014). Low risk and high return–affective attitudes and stock market expectations. *European Financial Management*, *20*(5), 995-1030.

L'Haridon, O., and Vieider, F. M. (2019). All over the map: A worldwide comparison of risk preferences. *Quantitative Economics*, *10*(1), 185-215.

Manski, C. F., and Neri, C. (2013). First-and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior*, *81*, 232-254.

Markowitz, H.M. (1952). *Portfolio Selection. The Journal of Finance.* ***7*** *(1): 77–91*

Menapace, L., Colson, G., and Raffaelli, R. (2015). Climate change beliefs and perceptions of agricultural risks: An application of the exchangeability method. *Global Environmental Change*, *35*, 70-81.

Murphy, A. H., and Winkler, R. L. (1974). Probability forecasts: A survey of National Weather Service forecasters. *Bulletin of the American Meteorological Society*, *55*(12), 1449-1452.

Nosić, A., and Weber, M. (2010). How riskily do I invest? The role of risk attitudes, risk perceptions, and overconfidence. *Decision Analysis*, *7*(3), 282-301.

Shefrin, H. (2001). Do investors expect higher returns from safer stocks than from riskier stocks?. *The Journal of Psychology and Financial Markets*, *2*(4), 176-181.

Shefrin, H., and Statman, M. (1995). Making sense of beta, size, and book-to-market. *Journal of Portfolio Management*, *21*(2), 26.

Sonsino, D., Lahav, Y., and Levkowitz, A. (2021). The conflicting links between forecast-confidence and trading propensity. *Journal of Behavioral Finance*, *22*(4), 443-460.

Sonsino, D., Lahav, Y., and Roth, Y. (2022). Reaching for Returns in Retail Structured Investment. *Management Science*, *68*(1), 466-486.

Sonsino, D., and Regev, E. (2013). Informational overconfidence in return prediction–More properties. *Journal of Economic Psychology*, *39*, 72-84.

Stoimenov, P. A., and Wilkens, S. (2005). Are structured products 'fairly'priced? An analysis of the German market for equity-linked instruments. *Journal of Banking and Finance*, *29*(12), 2971-2993.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*(1), 39-60.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. S*cience*, *185*(4157), 1124-1131.

Tversky, A., Kahneman, D., (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.

Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297-323.

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge university press.

Wallsten, T. S., Shlomi, Y., Nataf, C., and Tomlinson, T. (2016). Efficiently encoding and modeling subjective probability distributions for quantitative variables. *Decision*, *3*(3), 169-189.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In *Risk aversion in experiments*. Emerald Group Publishing Limited.

Zhu, D., Wang, Q., and Goddard, J. (2021). A new hedging hypothesis regarding prediction interval formation in stock price forecasting. *Journal of Forecasting*.