Journal of Economic Psychology 39 (2013) 72–84

Contents lists available at SciVerse ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Informational overconfidence in return prediction – More properties ☆

Doron Sonsino *, Eran Regev

COMAS, College of Management Academic Studies, Rishon LeZion, Israel

ARTICLE INFO

Article history:
Received 13 April 2013
Received in revised form 23 June 2013
Accepted 27 June 2013
Available online 4 July 2013

JEL classifications:
D84
C53
C93

PscyINFO classifications:
2240
2340

Keywords:
Informational overconfidence
Optimism
Perceived volatility
Anchoring and adjustment

ABSTRACT

A field experiment revealed 3 forms of unrealistic optimism in skilled investors' interval predictions of future stock returns. The judgmental intervals were about 50% shorter than realized spreads in recent 3–6 months histories, suggesting that "underestimation of volatility" persists past the financial crisis. The intervals, however, rapidly widened as predictions diverged from zero, and a complementary technical-forecasting experiment showed that the increased spread pattern emerges even when volatility is accounted. The results support "anchoring with noisy monotone adjustments" and suggest that overconfidence hazards may instinctively attenuate when expectations get extreme.

1. Introduction

The irrational belief in precision of private assessments is considered a primary flaw of individual judgment (Lichtenstein, Fischhoff, & Phillips, 1982). When decision makers provide 80–90% confidence intervals for unfamiliar quantities, in particular, the actual hit rate is frequently lower than 50% (e.g., Biais, Hilton, Mazurier, & Pouget, 2005; McKenzie, Liersch, & Yaniv, 2008; Soll & Klayman, 2004).[1] Recent studies demonstrate that irrationally narrow intervals also emerge in expert financial prediction (for a range of examples see Ben-David, Graham, & Harvey, 2010; Deaves, Lüder, & Schröder, 2010; Glaser & Weber, 2007; Oberlechner & Osler, 2011). The overconfidence in subjective beliefs has drawn substantial interest in finance

☆ The paper was presented at the 2012 ESA conference in New York, the Erasmus-Technion 2012 Deadsea conference, the 2011 Experimental Finance meetings in Innsbruck and the 4th (2011) psychology and Investment conference at the College of Management. We thank Yoram Eden, Ido Erev, Dan Friedman, Tommy Gärling, Martin Holmen, Kim Kaivanto, Oren Kaplan, Ran Lachman, Avi Messica, Stefan Palan, Charles Plott, Max Shifrin, Eyal Weinstein, Eldad Yechiam and 2 referees for constructive comments. We thank the research authority at COMAS for financial support.

* Corresponding author. Address: School of Business, The College of Management Academic Studies, 7 Rabin Blvd., P.O.B 9017, Rishon LeZion 75190, Israel. Tel.: +972 3 9634485.

E-mail address: sonsinod@colman.ac.il (D. Sonsino).

[1] Throughout the paper we use "hit rate" or "calibration rate" to denote the proportion of cases where the hidden quantities fall within the respective intervals (e.g., Soll & Klayman, 2004).

0167-4870/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.joep.2013.06.006

research. Theoretical models show that informational overconfidence may boost trading volumes, affect prices in competitive markets, and influence profitability (see Subrahmanyam, 2008 for specific references). Experiments confirm that calibration-based overconfidence (henceforth: CBO) positively correlates with individual propensity to trade (Deaves, Lüder, & Guo, 2009), while adversely affecting performance in experimental asset markets (Biais et al., 2005; Kirchler & Maciejovsky, 2002).[2]

In experiment I we employed the field-based experimental approach (Harrison & List, 2004) to explore the predictions of competent Tel-Aviv stock exchange investors for the highly volatile, past crisis market of 2010–2011. The prediction assignments were framed as incentivized professional consultation tasks and participants were requested to provide 95% confidence limits for the return that familiar stocks would show in the 3-months following consultation. The target stocks were randomly drawn for each participant to test predictions on a broad collection of leading stocks, and questionnaires were distributed by email, explaining that the 3-months countdown would start at the date where the completed form is returned. The 93 preregistered participants (mean age 33; 54% MBAs; 40% reporting investment-industry experience) could thus deliberate their predictions freely, and even time their exact 3-months test period. While we hypothesized that the ambiguous economic conditions would push calibration rates close to the 90% rational benchmark, the results revealed that field players still administer drastically unrealistic expectations. The eventual hit rate (across 930 intervals) was only 27.5% and the median was even lower at 20%. Realized return fell beneath the 5% low confidence limit in almost 60% of the cases, revealing a particularly strong form of unrealistic optimism (Weinstein, 1980) in short-run stock return forecasting. Using the stock-specific historical return series for each date of participation, we moreover demonstrate that calibration rates could double or triple if the participants were adapting to realized fluctuations in very recent 3–6 months histories. With all these respects, the data suggests that the financial crisis and strong subsequent uncertainty did not alleviate the fundamental tendency for informational overconfidence.

In addition, we exploit the large sample of semi-professional predictions, for closer look into the exact patterns of informational overconfidence in finance-related prediction. While theoretical finance papers are able to gain much insight assuming that overconfident traders constantly underestimate volatility (e.g., Daniel, Hirshleifer, & Subrahmanyam, 1998), we more closely hypothesized that confidence may naturally decrease with the extremity of predictions. The prediction intervals centered at 10%, for specific example, should be longer than the intervals with midpoint at 5%, conveying lower level of subjective confidence.[3] A similar pattern is intuitively expected on the negative side, with the forecast intervals increasing in length as the midpoint turns more negative. The field-based prediction data strongly supported the conjectured pattern. A median split, for example, revealed that the relatively extreme predictions were 50% longer than the less extreme forecasts (mean interval lengths of 18% vs. 12% respectively). The increase in length with absolute predictions reflects in the predictions for specific stocks, shows in regressions controlling for other determinants of predictions, and it is even confirmed in individual-level comparisons. We propose 2 formal explanations for the increased length pattern, beyond the intuitive appeal. The first follows from the observation that economic return processes are frequently heteroskedastic with the conditional variance increasing in recent squared or absolute realizations (cf. the GARCH formulation of Bollerslev, 1986). If the noise in return processes increases with the magnitude of returns, judgmental confidence may diminish in parallel, and confidence lengths would increase as the intervals diverge from zero for "statistical reasons". Alternatively, the increased length pattern could nicely fit Tversky and Kahneman's (1974) Anchoring and Adjustment Theory (AAT). Decision makers that comply with AAT derive their best estimate of the target quantity first, and adjust the anchor upwards and downwards to obtain the confidence bounds. As adjustments from the anchor are insufficient, the emerging intervals are generally too tight. If the adjustments, however, increase with the absolute value of the anchor, then the increased length pattern may follow from "technical reasons" beyond the statistical explanation.

To further explore the statistical and technical motives we devised a short in-class technical forecasting experiment, where finance students predicted the monthly return on anonymous stocks from only few statistics regarding performance in 12 preceding months. The return series for each prediction assignment were drawn from historical S&P500 records, but the identity of underlying stocks and exact dates of inspection were concealed to control the information that subjects may access (Sonsino & Shavit, in press). By including the standard deviation of monthly returns as one of only 6 statistics that subjects observe regarding each unidentified series, we control the perceived volatility of return in month 13, the target of prediction. The results of the secondary experiment confirmed that both statistical and technical motives play significant role in the automatic increase in spread with absolute predictions. The confidence intervals that subjects delivered significantly widened with historical volatility (in line with the statistical explanation), while positive significant correlation between absolute point predictions and interval lengths still emerged when volatility was accounted (in line with the technical explanation), but only for highly volatile series. At the appendix we demonstrate that an AAT model where adjustments are subjected to constant noise while marginally increasing with the absolute value of the anchor and also increasing with the perceived volatility of the target, nicely captures the experimental results.

The increased spread pattern documented in our two experiments may hold interesting practical and theoretical implications. Primarily, the pattern suggests that the hazards associated with informational overconfidence may instinctively

[2] Glaser and Weber (2007) reject CBO showing that other facets of overconfidence (better-than-average scores) significantly correlate with trading frequencies.

[3] To encourage independent deliberation of the confidence limits and decrease experimental load, we did not elicit point predictions in experiment I. Extremity is therefore measured by the midpoint distance from zero.

attenuate when market conditions or personal expectations turn extreme. Willingness to purchase, in particular, may incrementally decrease with expected returns, if the parallel increase in expected volatility counterbalances the expectation for higher return. Similar anomalies might emerge on the sell side. The possible implications are further discussed in the concluding Section 7. Section 2 extends the literature review delineating our three main hypotheses. Section 3 describes the field experiment; Section 4 presents the results; Section 5 discusses the increased spread pattern. The technical forecasting experiment is summarized in Section 6.

2. Main hypotheses[4]

The primary goal of our study was to test if the 2008 crisis and subsequent turbulence have turned the extreme-case predictions of qualified investors relatively realistic. In addition to standard tests of accuracy and calibration, we aimed at comparing the length of prediction intervals to empirical estimates, to test the perception of stock return volatility past the crisis. The few preceding financial CBO studies that run such comparisons contrast the judgmental intervals with formal, statistical measures of volatility. Ben-David et al. (2010), for instance, contrast the S&P500 predictions of top executives with formal volatility estimates such as the realized volatility in lengthy historical series, or the standard deviation implied from option prices. Oberlechner and Osler (2011) even use GARCH to estimate the (conditional) spread in US dollar exchange rates for the specific dates where the interval forecasts were collected.[5] Both studies conclude that the intervals submitted by professionals are significantly too short relatively to the statistical spread estimates. The formal statistical approach, however, is based on strong assumptions regarding the stochastic process of returns and long run stability of parameters. The application of such stationary analysis in the current context, for estimating the spread in return of specific stocks from the highly volatile series around the crisis, appears very restrictive. We therefore adopt a basically different, adaptive approach, using the latest running series of (3-months) returns to arbitrarily approximate the recently "realized spread" for each forecast. When the 90% spread, for example, is calculated from the return series for the 40 trading-days preceding participation, the "realized spread" is derived by subtracting the 3rd lowest observation from the 3rd highest. The calculations are applied for each participant and each stock separately, utilizing the latest data that respondents could actually explore to approximate the magnitude of recent fluctuations. If the participants properly adapt to recent turbulence, the judgmental spread estimates should approach the "realized spreads":

H1. The distance between high and low confidence limits would match the "realized spread" in recent histories, reflecting the attentiveness of participants to recent fluctuations.

Our next 2 hypotheses deal with closer examination of informational confidence patterns. Mathematical finance models utilize the "underestimation of volatility" to formally characterize overconfident investors. The tendency for informational overconfidence is frequently assumed as an exogenous predetermined characteristic of some traders with perceived volatility strictly lower than actual volatility measures (the literature is too large for direct survey; see Subrahmanyam, 2008). However, constant variance seldom characterizes economic processes in reality. The assumption that conditional volatility may vary with recent realized returns is an essential component of canonical models of dynamic return processes (Bollerslev, 1986; Engle, 1982). Macroeconomic surveys show that subjective uncertainty regarding inflation rates tends to increase with inflation forecasts (Golob, 1994), while experiments prove that the spread between high and low stock return predictions increases with the volatility of the series that subjects observe (Du & Budescu, 2007). Hypothesis H2 therefore suggests that prediction lengths would increase as predictions turn more extreme:

H2. The spread between low and high confidence limits would increase as the midpoint of the prediction interval diverges from zero, showing that judgmental confidence decreases with the extremity of predictions.

The discussion above proposed that the increased length pattern may follow from the heteroscedastic nature of individual expectations. Alternatively, the varying lengths nicely fit into an extension of Tversky and Kahneman's (1974) Anchoring and Adjustment Theory. Decision makers that comply with AAT derive their best estimate of the target quantity first, and adjust the anchor upwards and downwards to obtain their confidence bounds. The process of adjustment from external or internal (self-generated) anchors still attracts diverse research. Epley and Gilovich (2006) propose that adjustments from self generated anchors proceed in iterative manner. The plausibility of estimates is reexamined at the end of each round, and adjustments continue until a plausible estimate is reached. Janiszewski and Uy (2008), in this spirit, demonstrate that adjustments from precise anchors (e.g., 3.998) are smaller than adjustments from rounded anchors (4 exactly), concluding that the presentation format of the anchor may affect the subjective scale of adjustment. If the scale of adjustments, in addition, positively correlates with the absolute size of the anchor, then total adjustments may increase in parallel, generating the increased length pattern witnessed in our data. In this case, the increase in spread between high and low forecasts follows from technical properties of interval prediction, beyond the statistical explanation discussed above. Our secondary technical-

[4] To keep the paper concise, we discuss the 3 innovative hypotheses skipping some obvious hypotheses regarding choice rates, calibration and accuracy. The analysis addresses these other issues directly.

[5] The unconditional approach was first implemented by Glaser, Langer, and Weber (2005, 2012) that showed that the monthly stock/index predictions of professionals and students are significantly too narrow relatively to statistical volatility measures. Equality could not be rejected for weekly forecasts. The derivation of volatility estimates from option prices is infeasible for the stocks composing TA25.

forecasting experiment tested the technical explanation by tightly controlling the information that subjects receive regarding the prediction target, including recent realized volatility. Hypothesis H3 stipulates that the increased length pattern would reflect even when volatility is accounted:

H3. The decrease in confidence with extremity of predictions would show in technical-forecasting tasks, where subjects predict subsequent returns from only few statistics regarding the past performance of the stock, including realized volatility.

3. Experiment I – method and participants

Preliminary calls for participation were distributed in MBA classes, alumni lists and selected financial portals. Interested subjects were requested to send a short registration form to the experiment's email and those that met some competence thresholds received a personal questionnaire in return.[6] Participants were obliged to report their formal id, phone number, and provide a residence address for payout delivery. The DATE where each completed questionnaire arrived at our email was fixed as the starting point for evaluating performance.[7]

Each questionnaire consisted of 10 binary choice problems: 5 BUY and 5 SELL tasks. Participants were requested to assume the role of an investments consultant and chose the best stock for purchase or sale, assuming their decisions would be tested 3 months after delivery. The instructions asked the presumed consultants to approach each problem independently, underlining that one of the 10 tasks (''the designated problem'') would be randomly drawn to determine eligibility for consultation bonus. If choice in the designated problem turned out correct (i.e., the stock selected outperformed the alternative in BUY or underperformed the alternative in SELL), the participant received 100 New Israeli Shekel (about 30 US$) for successful consultation. The forecasting assignments were integrated with the choice problems, asking respondents to provide 95% confidence limits for the test-period return on their selected stocks (see Web supplement I for the translated script). The instructions forewarned that confidence estimation is considered a challenging task and briefly discussed the tradeoffs of too narrow or excessively wide forecasts. An extra 80 NIS bonus was promised to respondents that meet 3 undisclosed conditions for ''fair and accurate'' prediction.[8]

The stocks for each binary choice, in each questionnaire, were randomly drawn from TA25, the list of leading 25 stocks at the local exchange.[9] The stocks composing TA25 are largest in market capitalization and represent some of the most familiar companies in the economy. The random drawing was meant to test selection and prediction abilities on a broad, ecologically valid dataset (cf. Roe & Just, 2009). An alternative, nonrandom design could bias the results, as performance may depend on the relative difficulty of selected choice problems and idiosyncratic market trends. By restricting the referenced stocks to TA25, however, we construct very challenging choice assignments; e.g., separating the two largest commercial banks for 3-months investment.

In addition to the 10 core choice and prediction tasks subjects were requested to fill-in a short risk-preference assignment, rank their approval of 4 statements regarding market efficiency, and provide inclusive details regarding their education and experience. The first completed questionnaire was received in October 13th 2010. In April 21, 2011 we stopped advertising with 93 eligible respondents (19 female; 74 male). On average, the participants reported 16.5 years of formal education, which implies some level of graduate studies. The average AGE was 33.3 with 54% holding or pursuing MBA. About 40% reported professional investment-industry experience, but only 25 (of 93) were holding industry jobs at the time of participation. The average tenure of the experienced respondents was 3.6 years; the maximum was 10 years.

4. Experiment I – results

4.1. Choice tasks

The bottom line results for the choice tasks suggest that the participants could not effectively separate pairs of TA25 stocks for 3 months horizon. On average, stock selection was correct in only 4.8 problems along the 10 problems questionnaires. The distribution of correct choice rates (henceforth, CCR) is provided in Fig. 1. About 50% of the sample shows worse than chance performance with CCR $\leqslant$ 40%. Only 35% achieve CCR $\geqslant$ 60%. The hypothesis CCR = 50% could not be rejected in standard tests; e.g., $p = 0.17$ by Wilcoxon signed-rank test.[10] Our results, with this respect, are similar to Törngren

[6] We insisted on finance related academic studies (74% of the sample) unless the participant classified her familiarity with the local market or suitability for professional investment consultation at level $\geqslant$3 in 7 points scale. The median subjective FAMILIARITY/THEORY/PROFESSIONAL FIT ranks were all 5.

[7] Because of the random arrival of questionnaires, we control for DATE-related variables throughout the analysis. Recent trends (approximated by the change in TA25 in the 1–3 months preceding DATE) did not affect performance.

[8] The consultation bonus should incentivize subjects to select the stock that is more likely to "win" in each problem. The incentivization of interval predictions is more complex. Jose and Winkler (2009) propose a linear scoring rule that is incentive compatible under risk neutrality. We alternatively employed informal incentives to enhance motivation independently of risk preferences.

[9] Stocks were drawn without repetition in 77 questionnaires. The results are robust to removal of the 16 questionnaires where stocks were drawn with repetition.

[10] Since the pairs of stocks were randomly selected and test periods varied with DATE, we treat individual observations (hit rates, CCRs, etc.) as independent for the statistical tests. We use 2-tails significance levels throughout the paper; ρ always denotes the Spearman rank correlation.

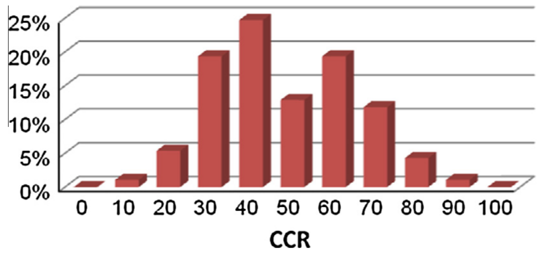


Fig. 1. Distribution of correct choice rate.

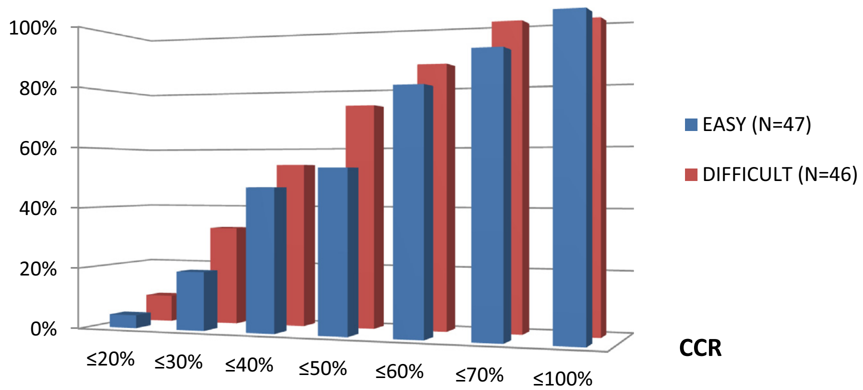


Fig. 2. CCR in difficult vs. easier problems.

and Montgomery (2004) where professionals showed CCR of 40% in binary stock selection for one month horizon. The failure in "picking the right stock" is robust and reemerges in various alternative tests; e.g., paired comparisons of the test-period return on selected stocks vs. non-selected stocks do not reveal stronger results for the stocks selected. The CCRs of professionals, MBAs, and subjects proclaiming strong familiarity were all close to 50%. The correct choice rates in BUY and SELL problems were almost equal but the correlation, more encouragingly, was positive $\rho = 0.15$.[11]

Closer analysis moreover revealed significant increase in CCR when the relatively difficult questionnaires are ignored. To approximate the difficulty of each choice, we use the absolute value of the difference in realized 3-months returns. If the 2 stocks composing a given choice assignment eventually showed drastically different results, then the problem is considered less difficult, assuming it was a priori easier to point at the best stock for purchase or sale. If, on the other hand, the test-period returns on the 2 stocks defining the problem were relatively close, the problem is considered difficult, assuming it was hard to separate the stocks ex-ante.[12] The difficulty of each questionnaire was approximated by averaging the difficulties of the 10 underlying problems and the average scores were normalized to $[0,1]$. DIFFICULTY shows significance in Probit regressions with model selection on the correct choice rates (see Web supplement II for illustrative results).[13] The mean CCR of the participants that confronted the most difficult questionnaires ($N = 46$; mean normalized DIFFICULTY 0.82) was 44%, compared to mean CCR of 51% to those that received the relatively easier versions ($N = 47$; mean DIFFICULTY 0.49) ($p < 0.05$ by Pitman test), and the easier questionnaires dominated the difficult in CCR distribution (Fig. 2). The mean CCR for the 5 most difficult questionnaires was 38% compared to 60% for the 5 easiest.

4.2. Return prediction

4.2.1. General

The 930 prediction intervals collected along the experiment were misplaced and far too short to accommodate the realized returns. The average hit rate was 27.5% and the median was even lower at 20%. Only 2 participants were perfectly calibrated with hit rate of 90%, while 2 subjects, with average interval lengths 3 times larger than others, obtained maximal hit rates of 100%. Our results, with this respect, compete with some of the worst CBO examples in the literature; e.g., Deaves

[11] The BUY and SELL dates were significantly different for only 3 stocks of 25. When the respective problems are removed, CCR stays 47.7% ($N = 813$).

[12] Lichtenstein et al. (1982) similarly use the ratio of hidden quantities to approximate the difficulty of almanac binary choice problems; e.g., the difficulty of choosing the most populated city is approximated by the ratio of larger to smaller populations. Correct choice rates decrease with difficulty.

[13] Before normalization, the difficulty scores lied between 4.51 and 16.48. Similar results emerged when DIFFICULTY was measured by the 3 months preceding the DATE of participation.

Table 1
Comparison of buy and sell predictions.

	BUY	SELL	Wilcoxon test
HIT RATE	24.1%	31.0%	$p < 0.01$
LGTH	12.5	11.1	$p = 0.02$
PRED	8.5	2.3	$p < 0.01$
%(PRED < 0)	4.5%	31.0%	$p < 0.01$
RETURN	−1.1	−1.8	$p = 0.05$
ERROR	9.6	4.1	$p < 0.01$
ABS(ERROR)	11.6	9.2	$p < 0.01$
LAST RETURN	5.4	4.1	$p = 0.10$
HIT RATE around LAST RETURN	33.1%	30.3%	$p = 0.32$

LGTH is the distance between the 95% and 5% confidence limits. PRED is the midpoint and % (PRED) < 0 denotes the proportion of intervals centered at negative return. RETURN is the realized return along the test period. ERROR = PRED-RETURN and ABS(ERROR) is the absolute value. LAST RETURN presents the return in the 3 months preceding DATE. HIT RATE around LAST RETURN is the hit rate of the shifted intervals with midpoint at LAST RETURN. The table presents the sample average for each condition. The Wilcoxon test is applied on individual differences ($N = 93$).

et al. (2009) 15–28% hit rates in 90% general knowledge confidence intervals. In the next paragraphs we separately explore accuracy and length to understand the roots of the extreme overconfidence.

4.2.2. Accuracy

While TA25 showed hesitant mixed trends in the months preceding participation, the prediction intervals were generally optimistic.[14] More than 95% of the BUY predictions were centered at positive return. The average midpoint (henceforth, PRED) was 8.5% ($N = 465$), revealing quite bullish expectations regarding the performance of stocks selected for purchase. While SELL predictions naturally appear much less optimistic, PRED was still positive 2.3% on average, with less than one third of the intervals centered at PRED < 0 (see Table 1). Realized returns, however, were mostly negative throughout the sample. The stocks selected for purchase decreased by 1.1% along the 3-months test periods, while the stocks selected for sale decreased by 1.8% ($p = 0.2$). On average, PRED was 9.6% higher than realized returns in BUY and 4.1% higher than realized returns in SELL. The very low hit rates therefore follow, at least partially, from the optimistic misplacement of predictions. Since predictions were relatively more realistic for the stocks selected for sale, SELL calibration rates were almost 7% higher ($p < 0.01$).

At the bottom of Table 1, we briefly examine if the participants could improve accuracy by placing their interval around the realized return at the 3 months preceding participation (henceforth: LAST RETURN). Paired comparisons of PRED and LAST RETURN suggest that the judgmental predictions were more optimistic than recent realized returns in BUY ($p = 0.06$), but equality could not be rejected for SELL ($p = 0.40$). When the intervals are symmetrically shifted from PRED to LAST RETRUN, the hit rate in BUY increases from 24.1% to 33.1%, while the hit rate in SELL stays 30.3%. The unrealistic optimism in BUY predictions could thus significantly reduce if subjects were placing their interval around LAST RETURN, but the accuracy of SELL intervals is similar to the accuracy of the shifted intervals.

4.2.3. Length

The optimistic misplacement of predictions would not have induced such low calibration rates if the intervals were long enough to compensate for the errors. The length of prediction intervals, in general, may serve as more precise measure of informational overconfidence (compared to hit rates), if a benchmark level of correct length could be defined. The studies of Ben-David et al. (2010) and Oberlechner and Osler (2011) tackle the challenge formally, contrasting the length of elicited intervals to formal, empirical spread estimates. Both studies conclude that the judgmental intervals are much too narrow relatively to various formal volatility estimates. The derivation of such formal spread estimates, however, frequently relies on lengthy series (e.g., Oberlechner and Osler use the sequences from 1970 or 1978 for the GARCH) and implicitly assumes that the stochastic process of return is stable. Since the current study deals with the forecasting of specific stock returns, past the extreme crisis turbulence, we choose to take a different approach using the recent return series for each stock to derive intuitive "realized spread" estimates that heuristically summarize the recent fluctuations. The calculations were conditionally implemented for each DATE and each TA25 stock separately, to exploit the most recent data that respondents could pursue to estimate volatility. Specifically, we used the overlapping 3-months return series for the 20, 40 or 60 trading days preceding participation to derive tentative spread estimates for each respondent. The calculations from 40 trading-days histories, for example, use the distance between the 3rd lowest and 3rd highest returns in the sequence to approximate the realized spread (see Appendix A for detailed example). The calculations for $T = 60$ similarly ignore the 3 most extreme observations, using the difference between the 4th lowest and highest returns, over the last 60 trading days, to estimate the 90% spread adaptively. The realized spread by the latest 20 trading days is similarly defined, using the 2-nd lowest and highest observations.

[14] About 62% of the completed questionnaires arrived after monthly decrease in TA25 (average decrease 2.49%); 38% arrived after monthly increase in the index (average 2.47%).

Table 2
Realized spread.

	Judgmental predictions	"Recent spread" estimates		
		$T = 40$	$T = 20$	$T = 60$
LGTH	11.8	18.0**	12.7**	21.5**
PRED	5.4	8.0**	6.7 (N.S.)	8.1**
ERROR	6.8	9.5**	8.2 (N.S.)	9.6**
HIT RATE	27.5%	40.5%**	34.7%**	46.0%**
HIT RATE of the extended intervals around PRED	NA	49.5%**	34.6%**	57.1%**

LGTH, PRED, ERROR and HIT RATE were calculated for the modified intervals similarly to the calculations for the elicited intervals. The extended intervals around PRED were derived by symmetrically extending (or contracting) each interval to match the realized spread. The asterisks summarize the results of paired comparisons between the modified and actual intervals ($N = 93$).

The results for $T = 40$ are summarized at the respective column of Table 2. The average realized spread is about 18.0 compared to average prediction length of 11.8 ($p < 0.01$). When the judgmental intervals are symmetrically extended (or contracted) to match the realized spread, the median hit rate increases from 20% to 50% while the hit rate climbs from 27.5% to 49.5%. A sign-test on individual hit rates confirms that the adaptive spread approximations could significantly increase calibration (hit rates increase for 63 participants while decreasing for only 30; $p < 0.01$). Qualitatively similar results emerge when the historical spread is calculated from extended series of 60 returns or shorter series of only 20 recent observations. The use of 60 observations to approximate the spread would have increased calibration to more than 57%. The median hit rate could triple from 20% to 60%. Interestingly, realized spreads are significantly longer than experimental predictions even when calculations are based on short series of only 20 overlapping returns, although the differences are small in magnitude. Our results therefore confirm, using an essentially different methodology, the tendency of skilled investors to underestimate volatility. The interval forecasts that our participants delivered seem to ignore the strong fluctuations in prices, as witnessed in most recent series. Our first main hypothesis H1 is therefore strongly rejected, suggesting that competent subjects are still unable to realistically grasp the volatility in stock-returns, in spite of the 2008 crisis experience.[15]

4.2.4. Cross-sample comparisons

The low calibration rates appear very robust and do not vary with competence, risk attitude and other personal attributes. The single constructive result from a comprehensive cross-sample analysis is a positive significant correlation between personal belief in professional stock selection and eventual calibration.[16] In the subsidiary tasks, respondents ranked their consent with each of the following statements in 1–5 scale: (C1) *Choosing a winning stock is like buying a lottery: picking a winning stock is a matter of luck* (C2) *Almost never do I buy or recommend disappointing stocks* (C3) *The stock market is rather efficient. Beating the market is extremely challenging*, and (C4) *I am frequently able to point at stocks that would outperform the market*. The results for C4 turned uninformative and the total (6-C1) + C2 + (6-C3) was selected as a personal CONTROL score. The higher CONTROL types delivered longer intervals (correlation of CONTROL with average interval LGTH $\rho = 0.29$) and achieved higher hit rates ($\rho = 0.26$). A median split revealed average LGTHs of 9.0 for the 52 subjects with lower than median CONTROL, compared to 15.4 for all others ($p < 0.01$). The respective HIT RATEs were 21.9% vs. 34.6% ($p < 0.01$). The HIT RATE of subjects with 10% highest CONTROL scores was 47.3%, almost twice larger than others. The high CONTROL participants were slightly older (average age 35.4 vs. 31.5; $p = 0.03$) but otherwise did not differ significantly from other respondents. Prediction errors, in real or absolute terms, were not affected by CONTROL scores (e.g., average absolute errors 10.1 vs. 10.5; $p = 0.63$). The observation that belief in professional stock selection negatively correlates with informational confidence seems intriguing as we are not aware of comparable evidence. Intuitively, the difference may follow from hidden motivational factors, as the participants that do not believe in professional stock-selection could casually deliver relatively tight intervals compared to the high control subjects that deliberated their predictions more carefully.

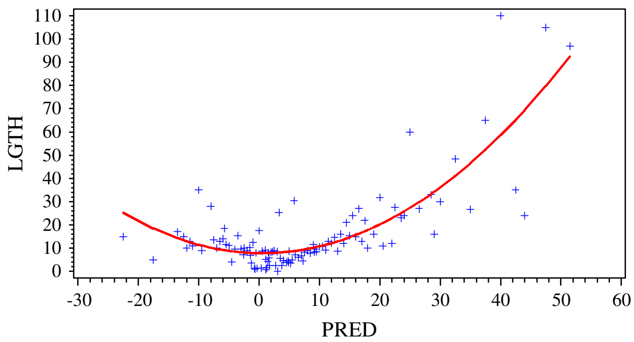
Finally note that stronger prediction abilities did not translate into superior stock selection in our sample. The correlation between HIT RATE and CCR was close to zero ($\rho = -0.07$) and even the participants with CCR > 60% showed about average calibration rate of 29%.

5. The decrease in confidence with absolute predictions

Hypothesis H2 is unequivocally supported by the data. The increase in interval lengths as the midpoint diverges from zero first emerges in a scatter plot of LGTH against PRED (Fig. 3). As many intervals cluster around similar midpoints, PRED takes

[15] To complement the adaptive analysis, we calculated the 90% spread from non-overlapping 3-months return series. Stocks with insufficient histories were ignored or classified by shorter histories. The empirical spread estimates strongly varied with the specific horizon assumed for the calculations, but were unequivocally larger than LGTH. Note also that the hit rate of the empirical confidence intervals were significantly lower than those of the extended intervals around PRED (except for $T = 20$); e.g., the HIT RATE of the $T = 40$ intervals was 40.5% compared to 49.5% for the extended intervals around PRED.

[16] Since prediction lengths increased with |PRED|, we account for PRED, PRED^2, etc. in the cross-sample regressions on LGTHs or HIT RATEs. Illustrative Probit results are provided in Web supplement III.



*The red line equation is 7.9-0.036*PRED+0.032*PRED^2

Fig. 3. Scatter plot of LGTH against PRED.

Table 3
Average LGTH for main PRED domains.

	PRED > 0	PRED < 0
0 < |PRED| < 5	6.9 (N = 233)	9.9 (N = 98)
5 ⩽ |PRED| < 10	8.0 (N = 271)	12.7 (N = 48)
|PRED| ⩾ 10	20.5 (N = 198)	24.3 (N = 19)

only 108 distinct values along the figure. Each point in Fig. 3 presents the average LGTH of all the intervals centered at the respective PRED. The figure clearly demonstrates that LGTH tends to increase with the absolute value of PRED. Nonlinear estimations reveal close to parabolic relation, suggesting that the increase in LGTH accelerates as PRED diverges from zero (the red line in the figure).[17] Table 3 discloses the average LGTH of the intervals centered at various PRED domains. Again, LGTHs increase with |PRED| and the rate of increase appears to accelerate. In addition the table suggests that LGTHs are larger where PRED < 0.

To compare the length of positive vs. negative predictions more precisely we construct a matched sample of all |PRED| values that admitted at least one interval centered at +|PRED| and another interval centered at −|PRED| and run a paired comparison of the average lengths. The intervals with negative midpoint are 28% longer than the intervals with positive midpoint (mean average lengths 12.4 vs. 9.7 respectively) and tests for equality of paired observations strongly confirm that predictions are longer on the negative side (e.g., sign-test $p < 0.01$). Closer look at the matched sample reveals that the proportion of intervals with positive upper limit among the predictions at PRED < 0 (33.5%) is twice larger than the proportion of intervals with negative lower limit among the predictions at PRED > 0 (16.4%; $p = 0.03$ by sign test on the paired proportions). The differences are much larger when the control for |PRED| is removed: the proportion of intervals with upper limit >0 amongst the intervals at PRED < 0 ($N = 165$) is 53.9% compared to 15.8% intervals with lower limit <0 amongst the predictions at PRED > 0. The longer intervals at PRED < 0 may therefore represent, in one more form, the persistent optimism of respondents: even when predictions are centered at the negative orthant, the upper limit is frequently positive, acknowledging the possibility of positive results.

The increase in LGTH as intervals diverge from zero shows in many forms of analysis. The Spearman coefficient of correlation between LGTH and |PRED| is 0.49 for the intervals with positive midpoint. The correlation is smaller, but positive 0.26 where PRED < 0. Positive significant correlations emerge for 22 of the 25 stocks in TA25 (in the PRED > 0 domain; samples are too small for stock-level analysis where PRED < 0). The pattern is robust to elimination of intervals with extreme lengths; it is confirmed in regression analysis (controlling for other predictors of LGTH such as CONTROL); and it even shows in individual level analysis. Details are provided in Web supplement V.

6. Experiment II

Hypothesis H3 conjectured that the increased spread pattern may still arise when the volatility of the target is accounted, suggesting that "anchoring and proportional adjustment" plays significant role in the derivation of interval return predictions.[18] To test the hypothesis we have run a technical forecasting experiment, asking subjects to forecast returns from only

[17] e.g., nonlinear least squares estimation of the equation $LGTH = a + b * PRED + c * PRED^d$ gave the results: $a = 7.9$, $b = -0.036$, $c = 0.033$, $d = 1.99$ ($p < 0.01$). The hypothesis $b = 0$ could not be rejected at $p < 0.1$; all other coefficients show significance at $p < 0.01$.

[18] The control on perceived volatility could be tested if we were able to collect subjective volatility estimates, independently of the elicited interval predictions. We believe that such volatility estimates would randomly cluster around the reported standard deviation (STD12) and would not correlate with LGTH. We did not implement this extra examination since direct elicitation of volatility estimates is problematic.

6 statistics regarding the past performance of the stock. The short questionnaire consisted of only 4 prediction tasks. Subjects received information on the return that an unidentified stock accumulated in the 12 months preceding prediction (MON1-MON12), the total return over the last 6 months (MON7-MON12), the standard deviation of monthly returns in 12 preceding months (STD12) and the realized return in each of the 3 most recent months (MON10, MON11 and MON12). The 6 statistics were presented in tabulated form (see Web supplement VI) and subjects filled-in a median forecast, as well as 5% and 95% confidence bounds for the return in MON13, the next month in the series. The instructions emphasized that each sequence was drawn from historical S&P500 records and MON13 was saved in our files to calibrate predictions. Subjects were discouraged from attempting to recognize the stocks from the statistics, explaining that one of the series could summarize the performance of an industry stock in the seventies while others refers to an Internet stock somewhere after 2000. The forecasting assignments were accordingly introduced as purely technical and the purpose of the questionnaire was described as "studying predictions from incomplete historical data".

Two of the 4 prediction tasks were framed as BUY problems while the other 2 were framed as SELL assignments. Subjects were requested to assume the role of investment consultants and advice their client whether to postpone the purchase or sale of the underlying stock by 1 month, from the end of MON12 to MON13. The questionnaire was distributed in elective finance seminars and subjects took 20–30 min to complete the 4 tasks.[19] The 4 historical series composing the prediction assignments were selectively sampled to introduce considerable variation in historical volatility. Two of the series were relatively stable with standard deviations 4.6 (Problem Low1) and 6.4 (Problem Low2). The other two series were much more volatile with standard deviations 12.5 (Problem High2) and 18.1 (Problem High1). We have run two versions of the questionnaire with samples of $N = 26$ and $N = 20$. The next paragraphs summarize the results concisely:

6.1. Predictions are not affected by the framing of problems as BUY or SELL

Sequences Low2 and High2 were presented in opposite BUY/SELL covers in the two versions of the questionnaire. Unsurprisingly, the framing did not affect the placement or length of the intervals (e.g., the median prediction for sequence High2 averaged at 4.6% in version 1 (BUY) compared to 3.6% in version 2 (SELL), $p = 0.61$. The average interval lengths were 22.2 and 21.1; $p = 0.80$).

6.2. Symmetry of |P50| and LGTH with respect to historical returns

The return sequences underlying problems High1 and Low1 were multiplied by -1 for version 2 of the questionnaire. Standard deviations were kept intact. The multiplication by -1 shifted the median predictions without changing the absolute value or the length of the intervals. The results are strikingly symmetric for sequence High1 where the median prediction averaged at +1.6 in version 1 compared to -1.6 in version 2 ($p = 0.16$). Absolute predictions and intervals lengths were very close; e.g., average LGTH 24.0 in version 1 compared to 27.4 in version 2 ($p = 0.52$). Similar, slightly weaker, symmetries emerged for sequence Low1 (see Web appendix).

6.3. Prediction lengths increase with historical standard deviation

Since the split into versions did not affect $|P50|$ and LGTH, we join the 2 versions for the remaining examinations. Table 4 clearly shows that prediction lengths increased with STD12 and a Kruskal–Wallis test confirms that the differences are significant at $p < 0.01$. Fixed effect regressions (controlling for heterogeneity in individual prediction lengths) suggested that STD12 is the only statistic showing significant consistent effect on LGTH across the sample. The intervals submitted for the series with highest historical volatility (High1; STD12 = 18.1), in particular, were 2.5 times longer, on average, from the intervals submitted for the series with lowest volatility (Low1; STD12 = 4.6). The intervals for High2 were longer than the intervals for Low2 in 40 of 46 cases. The increase in LGTH with STD12, however, was frequently disrupted in cases of relatively small differences in historical volatility: the intervals for Low2 were shorter than those for Low1 in 10 cases, while

Table 4
Main results of experiment II ($N = 46$).

Sequence	STD12	$\|P50\|$	LGTH	$\rho(\text{LGTH}, \|P50\|)$	LGTH/STD12
Sequence Low1	4.6	1.8	9.8	0.02	2.1
Sequence Low2	6.4	2.8	12.3	0.07	1.9
Sequence High2	12.5	6.3	21.7	0.30	1.7
Sequence High1	18.1	6.3	25.4	0.34	1.4

STD12 denotes the historical standard deviation for each sequence; $|P50|$ presents the absolute median prediction. The $|P50|$, LGTH, and LGTH/STD12 columns disclose the mean result for each sequence.

[19] Incentives were similar to Experiment I (see Supplement VI for details). The hit rate in Experiment II was 47.8% compared to 27.5% in Experiment I, in line with the suggestion that serially correlated cues may significantly improve calibration (Budescu & Du 2007).

the High2 intervals were longer than High1 in 17 cases. Intuitively, the violations may follow from secondary factors, beyond STD12, that affect predictions non-systematically and disrupt the ranking when volatilities are close. In Appendix B we model these nonsystematic, secondary factors as additive "white" noise.

The rightmost column of Table 4 still shows that, on average, the response to historical volatility exhibits standard patterns of diminishing sensitivity. When STD12 is only 4.6, the average prediction length is 2.1 times larger. The ratio LGTH/STD12 then constantly decreases with STD12 and the predictions for the most volatile series High1 are only 1.4 times larger than the respective STD12. In all 4 conditions, prediction lengths are much shorter than the 3.28 normal distribution benchmark.

6.4. Lengths increase with absolute median predictions, but only for high STD12 series

The increase in length as intervals diverge from zero that characterized the predictions in Experiment I, emerges only for the high volatility sequences in Experiment II (see the correlation column of Table 4). The correlation between $|P50|$ and LGTH is positive 0.34 ($p < 0.05$) for the sequence with highest STD12. A median split of the High1 intervals by absolute median predictions reveals an average length of approximately 20 for the intervals with $|P50| \leqslant 6.5$ compared to average length of 32 for the intervals with $|P50| > 6.5$ ($p < 0.04$). Similar, but weaker, results emerge for the second most volatile sequence High2, where the correlation between LGTH and $|P50|$ is 0.30 and a median split reveals average lengths of 18.6 (25.6) for the intervals with lower (higher) $|P50|$. The positive correlations however disappear for the low volatility series. A median split of the Low1 intervals, for example, reveals an average LGTH of 10.4 for the intervals with $|P50|$ closer to zero, compared to average LGTH of 9.2 for the other predictions ($p = 0.65$).[20] The correlation between LGTH and $|P50|$ in Low1 is only 0.02. The results for the second least volatile sequence Low2 are similarly insignificant. In Appendix B we demonstrate that an AAT model where adjustments are subjected to vanishing noise while marginally increasing with (I) the absolute value of the anchor and (II) the perceived volatility of the environment, nicely captures the patterns of correlation between absolute predictions and interval lengths witnessed in the experiments.[21] When STD12 is relatively large, the errors are too weak to distort the proportional adjustments and prediction lengths increase with the anchor. When volatility is low, the secondary affects get relatively stronger and the correlation dissolves.

7. Concluding discussion

Uncertainty about the distribution of returns, sometimes addressed as "model uncertainty", has been invoked as a major cause for low stock market participation (e.g., Cao, Wang, & Zhang, 2005). Malmendier and Nagel (2011) illustrate that young investors respond most fiercely to recent stock market movements; while Kuhnen and Knutson (2011) show that negative affect may decrease the willingness to take risk and damage confidence in investment decision. With this motivation, we expected that the recent financial turmoil would decrease the optimism and confidence of our relatively young respondents, pushing their extreme-case predictions to realistic levels. The results of experiment I still exposed extremely low calibration levels. In hindsight, the predictions appear unreasonably positive, reaffirming the tendency for groundless optimism in financial planning (Kahneman & Riepe, 1998). The unrealistic optimism manifested at 3 distinct levels: (I) eventual returns fell below the 5% confidence limits, in 50% of the SELL and 70% of the BUY predictions (II) even when participants submitted negative predictions, the upper limit of the interval was positive in 54% of the cases (III) BUY predictions were significantly more optimistic than realized returns in the 3-months preceding participation. The poor results, in terms of accuracy and calibration, suggest that optimism and self confidence may override the deterring affects of anxiety and uncertainty in case-specific investment decision. The extreme case predictions of our relatively young semi-professional participants appear disconnected from recent stock market realizations in spite of the strong past-crisis uncertainty.

The experiments, in addition, point at properties of finance-related informational overconfidence that were not yet documented in the literature that evolved on the topic. Analysis of stock return dynamics around the experiment, suggested that hit rates could significantly increase if subjects were adapting, intentionally or subconsciously, to the realized fluctuations in very recent histories. Moreover, the spread between high and low confidence bounds rapidly increased as predictions diverged from zero, suggesting that the hazards associated with informational overconfidence may self-correct when expectations turn extreme. The observation that expected returns and perceived transaction risk negatively correlate suggests, in particular, that the inclination of rational investors to trade stocks, long or short, may sometimes decrease with the expected return. The willingness to pay for a stock with conditional expected return 10%, for instance, may be lower than the willingness to pay for a stock with expected return 5%, if the estimated spread increases in parallel; especially when the investor is loss-averse or tends to overweight small loss-chance probabilities (Tversky & Kahneman, 1992). From informal descriptive standpoint, the heteroskedastic expectations-pattern documented in our experiments may explain the tendency of private investors to "stay on the fence" in times of strong market uncertainty. From a normative perspective, the results illustrate

[20] The respective average $|P50|$ values were 0.6 vs. 3; $p < 0.01$ by Pitman test. LGTH does not correlate with |P50| although the absolute median predictions show significant dispersion across the sample.

[21] The noise may represent the "secondary factors" that affect prediction lengths at the individual or task level. Noise plays a significant role in statistical overconfidence models such as Erev, Wallsten, and Budescu (1994), Juslin, Wennerholm, and Olsson (1999), Moore and Healy (2008) and others. Our extended AAT model, in contrast, does not aim at tracing the roots of overconfidence but describing a plausible AAT framework that fits the experimental results.

the importance of forewarning investors of the hazards brought by exaggerated private-signal confidence. Financial education programs are still scarce and it has been proposed that existing programs may boost self-confidence instead of alleviating the bias (Willis, 2008). On-line investment management generally increases the risks of biased information processing, but, on the opposite side, provides an opportunity for increased use of decision support tools that may reduce the dangers (Looney, Valacich, Todd, & Morris, 2006). If investors, for instance, are encouraged to simulate the distribution of return for different asset allocations (Kaufmann, Weber, & Haisley, 2013), their spread estimates may approach past realizations and the perception of risk and return would become more realistic.

Appendix A. Calculating the spread from historical series

For concrete example consider stock 2590248 by TASE formal indexing (www.tase.co.il). The stock was selected by id_S125 that delivered the questionnaire in January 19th 2011. To derive the experienced spread estimate for this case, we examine the sequence of returns between 25NOV2010 and 19JAN2011 (the 40 trading days preceding delivery). The calculations for this case are therefore based on the price series for 25AUG2010-19JAN2011. Along these dates the stock was traded at prices ranging from 188.4 to 275.1, a difference of more than 46% that explains the large spread estimate. The lowest 3-months return in the series was 9.38% (the price of the stock increased from 206.8 to 226.2 between 16SEP2010 and 16DEC2010). The highest return was 34.7% (price increase from 201 to 271.3 between 04OCT2010 and 04JAN2011). The third-lowest and third highest returns were 12.43% and 33.877% yielding a spread estimate of 21.44%. The confidence interval actually submitted by id_S125 for this stock ([3,10]) was 2/3 shorter and less optimistic than the empirical interval ([12.43, 33.88]). The realized return on the stock from 20JAN2011 to 20MAR2011 was negative −4.19%, which is by far out of the prediction range provided by the participant and also lies outside the longer empirical interval. When the experimental interval however is extended around the midpoint (6.5) to match the realized spread, the resulting interval [−4.22, 17.22] is just long enough to contain the realized return. When similar calculations are applied for the other 9 stocks selected by id_S125, the results suggest that the empirical method would have increased the length of prediction intervals by approximately 50%, modestly increasing the hit rate from 10% to 20%. Finally note that the LAST RETURN on 2590248 (the return from 19OCT2010 to 19JAN2011) was 24.6% which is almost 4 times larger than PRED. The extended interval around LAST RETURN [13.88, 35.32] is misplaced and too short to accommodate the realized return.

Appendix B. Anchoring with Noisy Monotone Adjustments

The appendix briefly extends Tversky and Kahneman's (1974) AAT to explain the pattern of correlations between prediction lengths and absolute predictions as observed in the experiments. We start from the space S of information signals upon which agents construct the confidence intervals. For simplicity, assume S is finite using $\mu(s) > 0$ to represent the probability of state $s \in S$. In the controlled information design of experiment II, for example, s may consist of the 6 statistics provided with each prediction assignment, but individual heterogeneity or cross-sample effects (e.g., range of CONTROL scores) may be accounted by extending the space.

An Anchoring with Noisy Monotone Adjustments (ANMA) process on S consists of the following elements:

(1) An anchor function. $A : S \to \mathcal{R}_+$ with $A(s)$ representing the anchor or point prediction from which the interval forecast is constructed when realized information is s. For convenience, we assume $A(s) > 0$ or every signal s, but the extension to negative anchors is straightforward.

(2) Similarity Partition. $\Pi(S) = \{S_1, S_2, S_3 \ldots S_m\}$ satisfying the standard conditions $S_i \cap S_j = \emptyset$ for each $i \neq j$ and $\cup_i S_i = S$, with each S_i representing a set of information signals that are similar in terms of perceived volatility of the prediction target. In experiment II, for example, signals may belong to the same similarity set when STD12 is similar. The conditional standard deviation $\sigma(A|\Pi(s))$ is used to represent the volatility of the prediction environment in state s, with $\Pi(s)$ representing the similarity set containing s. The environment is more volatile in s compared to s' when $\sigma(A|\Pi(s)) > \sigma(A|\Pi(s'))$. For the correlation analysis we assume that the anchor admits at least two distinct values in each set; i.e., for every $S_i \in \Pi(S)$, there are two states $s \in S_i$, $s' \in S_i$ such that $A(s) \neq A(s')$.

(3) Adjustment functions. $\Delta U: S \to R_+$ and $\Delta D: S \to R_+$ that represent the expected adjustments (before noise) around the anchor for each information signal s. $\Delta U(s)$ denotes the expected upward adjustment from $A(s)$, while $\Delta D(s)$ is the expected downward adjustment from the anchor. The expected confidence interval for information signal s is $[A(s) - \Delta D(s), A(s) + \Delta U(s)]$ and the expected prediction length is $\Delta U(s) + \Delta D(s)$. We henceforth abbreviate $ADJ(s) = \Delta U(s) + \Delta D(s)$ to represent the total expected adjustment. In addition, following the results of experiment II, we assume that expected adjustments increase with the anchor within similarity sets, while increasing with perceived volatility between similarity sets:

3.1. Adjustments increasing with A, within similarity sets

For every $S_i \in \Pi(S)$ and $s, s' \in S_i$ such that $A(s) > A(s')$, $ADJ(s) \geqslant ADJ(s')$.

3.2. Adjustments increase with Volatility, between similarity sets

For every $s, s' \in S$ such that $A(s) = A(s')$ and $\sigma(A|\Pi(s)) > \sigma(A|\Pi(s'))$, $ADJ(s) \geqslant ADJ(s')$.

The last component of the model is the noise in adjustments, as discussed with the results of Experiment II:

(4) Independent Adjustment Noise. Φ with mean zero and finite variance σ_Φ. The noise is fixed in distribution and independent of A. The length of predictions when noise is taken into consideration equals LGTH = $ADJ + \Phi$.

We now briefly demonstrate that ANMA may capture the pattern of correlation between prediction lengths and absolute predictions as observed in the experiments. First, it is easily verified that the correlation between LGTH and A, conditional on S_i, is positive:

By definition,

$$\rho(LGTH, A|S_i) = \frac{COV(LGTH, A|S_i)}{\sigma(LGTH|S_i) * \sigma(A|S_i)}$$

Since the adjustment noise Φ is independent of A,

$$\rho(LGTH, A|S_i) = \frac{COV(ADJ, A|S_i)}{\sigma(ADJ|S_i) * \sigma(A|S_i) + \sigma_\Phi * \sigma(A|S_i)}$$

Since ADJ monotonically increases with A on S_i (3.1), it follows from Schmidt (2003) that the covariance $COV(ADJ, A|S_i)$ is non-negative so that $\rho(LGTH, A|S_i) \geqslant 0$.

The basic ANMA framework however is too general to guarantee that the correlation increases as the environment becomes more volatile.[22] To explain the increased correlation in the simplest manner, we assume that adjustments are linear in the anchor within each similarity set; i.e., $\Delta U(s) = \alpha_U + \beta_U(s) * A(s)$ and $\Delta D(s) = \alpha_D + \beta_D(s) * A(s)$, where α_U, α_D are constants and $\beta_U{:}S \to R_+$ and $\beta_D{:}S \to R_+$ are $\Pi(S)$-measurable functions that represent the proportional response to the anchor within each similarity set. Assuming, in addition, that at least one anchor is common to each pair of similarity sets, 3.2 implies

$$(*) \quad [\beta_{Ui} + \beta_{Di}] > [\beta_{Uj} + \beta_{Dj}] \quad \text{whenever } \sigma(A|S_i) > \sigma(A|S_j),$$

where $[\beta_{Uk} + \beta_{Dk}]$ represent the fixed proportional adjustments on similarity set S_k.

With this additional assumption,

$$\rho(LGTH, A|S_i) = \frac{COV(ADJ, A|S_i)}{\sigma(ADJ|S_i) * \sigma(A|S_i) + \sigma_\Phi * \sigma(A|S_i)} = \frac{[\beta_{Ui} + \beta_{Di}] * \sigma(A|S_i)}{[\beta_{Ui} + \beta_{Di}] * \sigma(A|S_i) + \sigma_\Phi}$$

and $(*)$ implies that the correlation $\rho(LGTH, A|S_i)$ increases in $\sigma(A|S_i)$.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joep.2013.06.006.

References

Ben-David, I., Graham, J. R., & Harvey, C. R. (2010). *Managerial miscalibration. NBER working paper 16215*.

Biais, B., Hilton, D., Mazurier, K., & Pouget, S. (2005). Judgemental overconfidence, self-monitoring and trading performance in an experimental financial market. *Review of Economic Studies, 72*(2), 287–312.

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

Budescu, D. V., & Du, N. (2007). Coherence and consistency of investors' probability judgments. *Management Science, 53*(11), 1731–1744.

Cao, H. H., Wang, T., & Zhang, H. H. (2005). Model uncertainty, limited market participation and asset prices. *The Review of Financial Studies, 74*(4), 1219–1251.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance, 53*(6), 1839–1885.

Deaves, R., Lüder, E., & Guo, Y. L. (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance, 13*(3), 555–575.

Deaves, R., Lüder, E., & Schröder, M. (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior and Organization, 75*(3), 402–412.

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

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica, 50*(4), 987–1008.

Epley, N., & Gilovich, T. (2006). The anchoring and adjustment heuristic: Why the adjustments are insufficient? *Psychological Science, 17*(4), 311–318.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over and under confidence: The role of error in judgment process. *Psychological Review, 101*(3), 519–527.

Glaser, M., Langer, T., & Weber, M., (2005). *Overconfidence of professionals and lay men: Individual differences within and between tasks*?. University of Mannheim, working paper.

[22] See Web appendix VII for more detailed version of the model.

Glaser, M., Langer, T., & Weber, M. (2012). *True* overconfidence in interval estimates: Evidence based on new measures of miscalibration. *Journal of Behavioral Decision Making*.
Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review, 32*(1), 1–36.
Golob, J. E. (1994). Does inflation uncertainty increase with inflation? *Economic Review, 79*(3), 27–38.
Harrison, W. G., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.
Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science, 19*(2), 121–127.
Jose, V. R. R., & Winkler, R. L. (2009). Evaluating quantile assessments. *Operations Research, 57*(5), 1287–1297.
Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1038–1052.
Kahneman, D., & Riepe, M. W. (1998). Aspects of investor psychology. *Journal of Portfolio Management, 24*(4), 52–65.
Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science, 59*(2), 323–340.
Kirchler, E., & Maciejovsky, B. (2002). Simultaneous over- and under-confidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty, 25*(1), 65–85.
Kuhnen, C. M., & Knutson, B. (2011). The influence of affect on beliefs, preferences and financial decisions. *Journal of Financial and Quantitative Analysis, 46*(3), 605–626.
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge Univ. Press.
Looney, C. A., Valacich, J. S., Todd, P. A., & Morris, M. G. (2006). Paradoxes of online investing: Testing the influence of technology on user expectancies. *Decision Sciences, 37*(2), 205–246.
Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics, 126*(1), 373–416.
McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval predictions? What does expertise buy you? *Organizational Behavior and Human Decision Processes, 107*(2), 179–191.
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517.
Oberlechner, T., & Osler, C. (2011). Survival of overconfidence in currency markets. *Journal of Financial and Quantitative Analysis, 47*(1), 91–113.
Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics, 91*(5), 1266–1271.
Schmidt, K. D. (2003). *On the covariance of monotone functions of a random variable, mimeo*. Dresden University.
Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, Cognition, 30*(2), 299–314.
Sonsino, D., & Shavit, T. (in press). Return prediction and stock selection from unidentified historical data. *Quantitative Finance*.
Subrahmanyam, A. (2008). Behavioral finance: A review and synthesis. *European Financial Management, 14*(1), 12–29.
Törngren, G., & Montgomery, H. (2004). Worse than chance? Performance and confidence among professionals and laypeople in the stock market. *Journal of Behavioral Finance, 5*(3), 148–153.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.
Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806–820.
Willis, L. E. (2008). *Against financial literacy education*.