The conflicting links between forecast-confidence and trading propensity

Doron Sonsino, Yaron Lahav, Amir Levkowitz*

While finance studies suggest that forecast-confidence motivates trading, the experimental findings regarding the confidence-trading links are inconclusive and statistically weak. Attempting to bridge the gap, we modify the standard interval forecasting task to measure forecast-confidence more directly. The adapted task is utilized to test the confidence-trading correlations at the attitudinal level and in specific scenarios. The attitudinal test surprisingly reveals that forecast-confidence negatively correlates with the inclination to churn one's stock portfolio, although confidence in profitability indeed boosts the willingness to trade particular stocks. The attitudinal correlation is endogenous, brought by opposite personality and competence effects on confidence and trading.

Keywords: Forecast-confidence, trading propensity, personality, overprecision.

JEL classifications: D8, C9, G1

---

* The paper is forthcoming at the *Journal of Behavioral Finance*. Doron Sonsino (the corresponding author; email: sonsino.doron@gmail.co.il) is at the College of Law and Business (CLB), Ramat Gan, Israel and adjunct at the Economics department of Ben-Gurion University. Yaron Lahav and Amir Levkowitz are from the Guilford Glazer Faculty of Business and Management at Ben-Gurion University. We thank Ido Erev, the Technion R&D foundation, and the research authority at COMAS (College of Management Academic Studies), for supporting this research. The study was presented at the 2017 SPUDM conference at the Technion, the IAREP 2017 meetings in Israel, the 2017 European ESA meetings in Vienna and seminars at the Gothenburg Research Institute, the Technion, and Ben-Gurion University. We have benefited from exchanges with Klajdi Bregu, David Budescu, Ido Erev, Yoav Ganzach, Sven Hemlin, Magnus Jansson, Ido Kallir, Doron Kliger, Ferdinand Langnickel, Christoph Merkle, Orkan Nadirov, Nigel Harvey, Dilek Onkal and Eldad Yechiam.

1. Introduction

Plain intuition and finance theory propose that forecast-confidence may stimulate trading.

The intuition is easily illustrated. If two comparable investors expect the same excess return on a given stock, but one is almost certain that the stock would beat the benchmark while the other hesitates, the willingness to invest of the confident must be stronger. The argument still applies when investors disagree in expectations. When a prospective investor is confident that the stock would bring positive return while a holder of the stock is almost certain that the return will be negative, the likelihood of exchange climbs compared to the case where the potential trade participants are uncertain.

Finance studies rarely address forecast-confidence directly, but relatedly explore the impact of excessive confidence and overprecision. Overprecision, defined as the tendency to trust personal beliefs excessively, is one of three facets of overconfidence by Moore and Healy's (2008) taxonomy.[1] Exaggerated forecast-confidence classifies as a special case. Theoretical finance papers model the over-precise traders as types that underestimate market volatilities (Daniel et al., 1998) or overweight private signals (Gervais and Odean, 2001). Diverse models illustrate that overprecision may explain the persistence of non-profitable trading and shed light on empirical puzzles such as the short run momentum in stock returns (Daniel and Hirshleifer, 2015).

Given the finance literature interest and the appealing intuition it is puzzling to note that experimental studies fail to validate the forecast-confidence and excessive trading links. Experiments and surveys typically use interval-production tasks to test for overprecision. The respondents, for example, provide 80% confidence intervals for the S&P500 annual return and calibration is tested by comparing the hit rate of the intervals to the preassigned 80% confidence level (Ben-David et al., 2013). Overprecision is tested by deriving formal measures of forecast-

[1] Confidence and overconfidence positively correlate (see, for instance, the classic calibration curves in Lichtenstein et al., 1982) but still represent distinct constructs. Moore and Healy (2008) separate between *overestimation* of individual abilities, *overplacement* relatively to others, and *overprecision* in subjective beliefs. Several papers show that trading propensity may increase with overplacement or overestimation (Glaser and Weber, 2007; Michailova and Schmidt, 2016; Bergu, 2020), but we presently focus on forecast-certainty and overprecision. The terms *forecast-confidence* and *forecast-certainty* are used interchangeably.

overconfidence from the submitted intervals (Glaser et al., 2013). While the overprecision hypothesis is commonly supported, the results in terms of linking forecast-overprecision and trading are disappointing. The overprecision measures derived from interval forecasts do not correlate with empirical trading volume records (Glaser and Weber, 2007; Merkle, 2017) and also fail to correlate with trading decisions in experimental studies (Fellner-Röhling and Krügel, 2014; Broihanne et al., 2014).[2] The gap between the intuitively appealing theoretical results and the weak experimental findings motivates the alternative approach of the present paper.

Basically, we depart from preceding experimental studies of the confidence-trading links by switching from interval-production tasks to comparable interval-evaluation assignments (Winman et al., 2004). Instead of providing confidence intervals for the target return, the subjects submit a point forecast and assess the likelihood of return falling within a fixed distance (plus/minus δ) from their point estimate. Essentially, the subjects estimate the likelihood of showing small prediction errors when forecasting the returns on familiar stocks, but the task is phrased in neutral language asking subjects to estimate the likelihood of the given length interval around their point prediction. The forecast-overprecision hypothesis is tested by comparing the likelihood that subjects assign to small errors to the realized small-error rate.

Motivated by the results of particular confidence-trading studies (Nosic and Weber, 2010; Fellner-Röhling and Krügel, 2014), we also draw a distinction between attitudinal (trait-level) trading inclination and the willingness to buy or sell particular stocks in specific scenarios. A trading propensity questionnaire is developed to generally measure the inclination to churn one's stock portfolio (Dorn and Huberman, 2005). Incentivized framed-field (Harrison and List, 2004) stock purchase and sell assignments are built to test the willingness to trade in specific scenarios.

While we designed the study with the intent of closing an apparent gap in the confidence-trading literature, the results revealed that the forecast-confidence and trading links are more complex than hypothesized. Two pilot studies and the key experiment showed that at the general trait-level the tendency for increased forecast-confidence associates with smaller trading propensity, although confidence-in-profitability motivates case-specific trading as conjectured. The negative trait-like

[2] Nosic and Weber (2010; discussed later) is an exception.

correlation is endogenous, representing the tendency of forecast-confident types to alter their stock holdings less frequently. Cross-sample analysis pointed at key personality traits (Goldberg, 1993), such as conscientiousness, openness to experience, and neuroticism, that boost trading propensity while negatively affecting forecast-confidence. The confidence effect on the case-specific purchase or sale decisions, on the other hand, is evidently positive. Confidence-in-profitability, defined as the likelihood that the subject assigns to positive return, even shows stronger predictive power for the amounts that subjects decide to buy from particular stocks, compared to the return forecasts alone. Testing the interaction between the contradictory trait-level and case-specific correlations, we discover that highly confident types, in terms of generally exhibiting strong confidence in their forecasts, show 2/3 smaller, but still positive, responsiveness to confidence-in-profitability in their case-specific trading decisions.

Finance overconfidence studies repeatedly point at the hazards brought by exaggerated confidence (Barber and Odean, 2000; Barber et al., 2008; Forman and Horton, 2019). The current trait-level results however propose that forecast-confidence may prevent the losses that excessive trading brings. With this respect, our results relate to a conflicting stream in the literature, suggesting that self-confidence, and even overconfidence, associates with advantageous trading skills (Hirshleifer and Luo, 2001; Oberlechner and Osler, 2012). The results also contribute to the increased research on personality effects in financial decision (e.g., Becker et al., 2012; Rustichini et al., 2016), additionally suggesting that psychometric tests of personality may prove useful in financial advising (Tauni et al., 2017).

The next section of the paper surveys the experimental confidence-trading literature more closely, and the modified forecast-certainty task is introduced in section 3. The subsequent sections of the paper present our evidence for conflicting links between forecast-confidence and trading. Section 4 outlines the experimental design and sections 5-8 present and discuss the results. Since our results regarding forecast-overconfidence and trading were not informative, we relegate the discussion of overprecision and trading to Section 9 of the paper. Section 10 concludes.

2. Literature review

Experimental studies use interval-production tasks to test for overprecision. In a common format of the assignment, the confidence level CONF is exogenously provided and the subject submits a lower limit F_{min} and an upper limit F_{max} such that the resultant $[F_{min}, F_{max}]$ interval is expected to accommodate the target with likelihood CONF. Judgmental psychology studies employ the interval-production methodology to elicit confidence intervals for unknown (hidden) quantities such as the population of a given city (e.g., Lichtenstein et al., 1982). Currently, we focus on finance contexts, where the prediction target is the future return R of some stock or index, as in Figure 1.

Figure 1: The Standard Interval Forecasting Task

This assignment deals with forecasting the return on *StockName* in October-December, 2016 (the last quarter of 2016):

-Submit a median prediction for the return on *StockName* in the last quarter of 2016 ____

- With probability 95%, I believe that the return on *StockName* in the last quarter of 2016 will be smaller than ________

- With probability 95%, I believe that the return on *StockName* in the last quarter of 2016 will be larger than _________

The subjects are faced with several interval-production assignments, so that calibration can be tested at the individual level. The term HIT is used for cases where the realized *r* (or the hidden quantity) falls within the interval; a MISS occurs otherwise. The responder is classified as *perfectly calibrated* when the hypothesis (HIT rate = CONF) cannot be rejected. Judgmental psychology experiments, however, consistently reveal hit rates significantly lower than the preassigned CONF (Moore and Healy, 2008), and return forecasting studies expose especially poor calibrations. The hit rates are more than half smaller than expected in the return forecasting studies of Ben-David et al. (2013), Sonsino and Regev (2013), Langnickel and Zeisberger (2016) and many others.

Measures of overprecision are derived from confidence intervals in two distinct methods. The next paragraphs discuss the two approaches, separating between *miscalibration-base*d and *volatility-based* measures of overprecision (Glaser et al., 2013).

<u>Miscalibration-based overprecision</u>: Diverse studies use the miscalibration score, calculated by subtracting the hit rate from the benchmark confidence level, as the overprecision metric. If, for example, the hit rate of 90% confidence intervals is only 50%, the 40% gap between the preassigned (expected) rate and the realized rate represents excessive confidence in subjective beliefs, in line with overprecision. Miscalibration rates, however, jointly depend on the accuracy of the intervals and their length (Soll and Klayman, 2004). Low calibration may be brought by largely inaccurate predictions, too tight intervals or a combination of both. The joint effects are easily illustrated using a return forecasting example. If the target return R_{t+1} is normally distributed around μ_t with standard deviation σ_t, then underestimation of σ_t by 25% would shorten the confidence interval, decreasing the hit rate to 78%, while 50% discount of σ_t would cut the hit rate further to 59%. It is easily verified, however, that hit rates of 78% (59%) alternatively emerge when the point estimate is 0.84 (1.4) standard deviations from the true mean. The accuracy-length tradeoff clearly shows in McKenzie et al. (2008), where IT experts provide more accurate, but shorter, prediction intervals for quantities related to their expertise. The experts' miscalibration rates are similar to those of non-expert students, although their shorter intervals apparently point at stronger confidence.

Moreover, the results of experimental tests of the miscalibration and trading links are typically negative. Glaser and Weber (2007) find that the trading volumes and portfolio turnovers of German investors increase with their belief in holding better-than-average investment skills, but fail to correlate with their miscalibration scores. Similar conclusions emerge in a more recent analysis of Barclays Bank trading records (Merkle, 2017). The miscalibration and trading-volume link is also rejected in asset market experiments (Biais et al., 2005; Fellner-Röhling and Krügel, 2014).[3]

[3] In Biais et al. (2005) miscalibration links with lower earnings in experimental markets. The current study is not designed to test the forecast-overprecision effect on earnings from trade (see the discussions in sections 9-10).

Volatility-based overprecision: As an alternative approach, return forecasting studies use the elicited intervals to estimate the *perceived volatility* of the target return. The perceived volatility estimates are contrasted with *empirical volatility* estimates and the difference (empirical volatility minus perceived volatility) is utilized as the forecast-overprecision measure.[4] Again, the overprecision hypothesis is typically supported (Graham and Harvey, 2001; Oberlechner and Osler, 2012; Ben-David et al., 2013). In Merkle (2017), for example, the perceived volatility of the quarterly FTSE return is about 50% smaller than historical volatility estimates.

The derivation of perceived and empirical volatility estimates, however, rests on strong statistical assumptions regarding the dynamic process of the returns (Sonsino and Regev, 2013). Moreover, empirical studies of the link between volatility-based overprecision and trading propensity again show inconclusive results. In Glaser and Weber (2007) volatility-based measures of overprecision fail to correlate with empirical trading records. Nosic and Weber (2010) encouragingly show that the willingness to invest in five German DAX stocks increases with volatility-based overprecision, even when forecast-optimism and financial risk tolerance are controlled in multivariate analysis. These positive results, however, disappear in Broihanne et al. (2014) although the two studies employ a similar design, and the more recent analysis of Merkle (2017) oppositely finds weak negative links between volatility-based overprecision and the trading frequencies of UK investors.

A distinct (third) approach to testing the overprecision-trading correlation is adopted by Fellner-Röhling and Krügel (2014) who find positive correlation between the overweighting of private signals in a 60-round signal detection task and trading volumes in unrelated experimental markets. The correlation, however, is statistically weak and only shows in one of two experimental conditions. While the Nosic and Weber (2010) results for the five DAX stocks show that confidence and willingness to invest may positively correlate at the case-specific level, the Fellner-Röhling and Krügel (2014) findings propose that confidence and trading propensity may link at the more basic trait-level. At the next section we introduce an alternative return forecasting task, aimed at alleviating the concerns that arise when using standard intervals to derive forecast-

[4] We adopt the Merkle (2017) definition of volatility-based overprecision. Other studies use slightly different definitions. Nosic and Weber (NW; 2010), for example, use the negative (-perceived volatility)/(empirical volatility) ratio. When the prediction target is fixed across the sample (as in NW), the empirical volatility estimates are fixed, so differences in overprecision only follow from differences in perceived volatilities.

certainty and overprecision metrics. The modified task is utilized to test the confidence-trading link at the two levels: (a) general trait, checking if subjects that show stronger confidence in their point forecasts also exhibit stronger inclination to trade stocks in an attitudinal survey developed for the current study; and (b) case-specific, testing if forecast-confidence affects the amounts that subjects choose to buy or sell from given stocks, controlling the subjective return expectations.

Figure 2: The Forecast Accuracy Assessment Task (FAAT)

-Submit a median prediction for the return on *StockName* in October-December, 2016 (the last quarter of 2016) _______

-What, in your opinion, is the probability that the return on *StockName* in the last quarter of 2016 would fall in a range of plus or minus 5% from the median?

The next diagram may help you develop your estimate:

[lower bound_____________median forecast ______________upper bound]

Add 5% to the median prediction and fill in the "upper bound" box.
Subtract 5% from the median prediction and fill in the "lower bound" box.

Submit the probability you assign to quarterly October-December, 2016 *StockName* return within the interval between the lower and upper bounds (i.e., in a range of plus or minus 5% from your median forecast): _______ (between 0% and 100%)

3. The forecast accuracy assessment task

The modified task, addressed as FAAT for forecast accuracy assessment task, consists of three steps (see Figure 2).

1: The subject submits a median forecast F for the target return R. The instructions explain that a median forecast is a positive or negative point estimate such that the provider assigns equal 50% likelihoods to larger or smaller return.

2: The instructions guide the subject to construct a fixed length interval around F, adding and subtracting a given margin δ from the median.

3: The subject provides a likelihood assessment CONF for the interval $[\mathrm{F}-\delta,\ \mathrm{F}+\delta]$, estimating the likelihood of return falling within the 2δ interval centered at F. If, for example, the median forecast is 7% and the margin δ is 5%, then the subject estimates the likelihood of the [2%, 12%] interval. Likelihood assessments can take any value between 0% and 100%.

As in the experiments with interval-productions, the subjects are faced with a sequence of such FAATs and the HIT rate is defined in the usual way. The difference between the average likelihood that the subject assigns to calibration and the realized hit rate is used as the overconfidence measure: $OC = \text{average}(CONF) - HIT\ rate$. If the CONFs, for example, are 70%, 50%, 90%, 70% while the realized return falls within the interval in only two of the four cases, OC is 20%. The subject overestimates the hit rate by 20%, in line with the overprecision hypothesis. [5]

While FAAT and the standard interval forecasting task may appear quite similar, the tasks are fundamentally different. In particular, FAAT improves on the standard interval forecasting task by four aspects:

(1) In the modified task, a HIT occurs when $\mathrm{F}-\delta \leq r \leq \mathrm{F}+\delta$; i.e., when $r-\delta \leq \mathrm{F} \leq r+\delta$. The hit rate of FAATs therefore only depends on the accuracy of the median forecasts. As overconfidence is measured by the difference between the average CONF and HIT, the impact of accuracy and confidence on overprecision can be separately assessed. An overconfidence score of 50%, for example, can emerge when an 100% confident subject (average CONF=100%) shows hit rate of 50%, or alternatively arise when a subject with confidence 75% shows half smaller hit rate of 25%. Subjects can be classified as relatively overconfident for showing stronger confidence than others, for exhibiting lower accuracy, or for both reasons. Such clean separation is impossible when miscalibration is measured using standard interval-production.

(2) By subtracting the realized HIT rate from the average CONF, FAAT defines an intuitively appealing overprecision measure representing the difference between the likelihood that subjects

[5] Interval-evaluation tasks were applied in few judgmental psychology studies (Murphy and Winkler, 1974; Winman et al., 2004; Teigen and Jørgensen, 2005; Speirs-Bridge et al., 2010), but we are not aware of financial forecasting studies employing a similar method. More detailed discussion of FAAT and the background judgmental psychology literature is provided in Sonsino et al. (2020).

assign to having smaller than δ prediction errors and the actual frequency of such small errors. Judgmental psychology studies propose that assessing the likelihood of given events is easier and more natural than making quantile predictions. Abbas et al. (2008), for instance, compare the estimation of beta distributions using probabilistic (likelihood) assessments to the estimation of the same distributions using quantile assessments. The method building on likelihood assessments is preferred by most subjects and ranks higher in monotonicity. Similar results emerge in Wallsten et al. (2016).

(3) Likelihood assessments can be easily incentivized using quadratic scoring rules. The incentivization of quantile assessments is more complicated. [6]

(4) Compared to volatility-based overprecision measures, FAAT bypasses the statistical problems that arise with deriving perceived and empirical volatility estimates.

4. The experiment

4.1: Procedure

The data was collected in MBA finance classes within four days in early August 2016, and the participation time was not effectively constrained. The instructions (Web supplement 1) were presented verbally and subjects received a printed version. Surfing the Web, public comments, or page turning were forbidden. The tasks were divided between three questionnaires: Q1-Q3 (Web supplements 2-4). Q1 consisted of eight FAATs and four case-specific BUY/SELL assignments. Q2 presented the trading propensity questionnaire and a sociodemographics survey. Q3 consisted of five short personality tests including a 44 problems big five. The questionnaires were distributed in fixed order, collecting the preceding booklet before distributing the next. The FAATs and BUY/SELL assignments were incentivized. The incentivization method is described next, along

[6] A loss function for the α quantile of X is $|x-q|\cdot(\alpha\cdot 1_{\{x>q\}}+(1-\alpha)\cdot 1_{\{x\leq q\}})$, where q is the elicited α quantile and x is the realized X, while a quadratic scoring rule for the likelihood p of event E is $p^2\cdot 1_E+(1-p)^2\cdot 1_{E^c}$ (Gneiting and Raftery, 2007). The quadratic score takes only 2 values, while the loss function for α varies with x. Details on the incentivization method are provided in Section 4.2.

the detailed description of the tasks. The next sections report the results for the N=72 MBAs (60% males) that completed the FAATs and the trading assignments with no errors or omissions. [7]

4.2: The FAATs

The prediction targets for the eight FAATs were stocks from TA25, the list of 25 leading stocks of the Tel-Aviv exchange (see Web supplement 5 for details). The prediction period was the last quarter of 2016. The first four tasks dealt with particularly familiar stocks and did not provide information regarding the past performance of the targets. Two tasks used 5% margins around the median. The other two tasks used 10% margins. The next four FAATs were similar in structure except for being preceded by charts of the daily price trends and trading volumes in the first six months of 2016 (see the appendix for an example). Again, two tasks used 5% margins, while the other used 10%. The abbreviations H (for history) and NH (for no history) henceforth separate the two types of tasks. The indices 10 and 5 similarly distinguish the 10% and 5% tasks. The symbols $CONF_{10}$, $CONF_{5}$, $CONF_{H}$ and $CONF_{NH}$ accordingly denote the average 10%, 5%, H and NH confidence levels. Unindexed CONF represents the task specific or the average confidence across the eight tasks, depending on context.

To incentivize the FAATs we used random task selection combined with a binarized scoring rule procedure. The random task selection is meant to motivate subjects for independent work in each task (Cubitt et al., 1998; Hey and Lee, 2005). Binarized scoring rules are used to prohibit the bias that personal risk attitudes may impose on experimental responses (Hossain and Okui, 2013; Harrison et al., 2014). In particular, the instructions explained that one of the forecasting or likelihood assessment tasks would be randomly selected to determine eligibility for a fixed 100 NIS payoff (about 26 US$). The probability of winning the 100 NIS decreased with the absolute error of the median forecasts and increased with the likelihood that the subject assigned to the realized events (hit or miss). The payment assignment was randomly drawn at the end of each session and used in January 2017 to determine eligibility for the 100 NIS. The students were invited to supervise the process and keep record of the draws.[8]

[7] Similar results emerge when the sample is extended to N=93 subjects with 1-2 errors or omissions and the problematic tasks are ignored. The results for the extended sample are presented in Web supplement 6.

[8] The probability of winning the 100 NIS was 100% for prediction errors (|F-r|)) smaller than 1%, 98% for errors smaller than 2%, etc. The quadratic scoring rule $100*[1-(1-P)^2]$, where P represents the likelihood assigned to the

4.3: The BUY/SELL tasks

To collect case-specific trading decisions, we add BUY/SELL assignments to the four FAATs where subjects received historical charts. In the two BUY problems, subjects were requested to assume they hold a 100,000 NIS deposit that would pay 2% quarterly return at the test period, but have a possibility to shift some amount to the prediction stock. The instructions explained that the *final portfolio return* would be derived from the BUY amount and the realized return *r*, by the formula BUY**r*%+(100,000-BUY)*2%, asking subjects to decide how much of the 100,000 deposit they choose to shift to the stock. The two SELL problems were similarly designed. The subjects were asked to assume that they currently hold 100,000 NIS of the stock, but have a possibility to liquidate any amount for risk-free investment at a quarterly rate of 0.5%.[9] The final portfolio return was derived by the formula SELL*0.5%+(100000-SELL)**r*%. For each BUY/SELL task, three participants were randomly selected to receive 1% of their final portfolio return.

4.4: Attitudinal trading propensity

The trading propensity questionnaire was developed in two pilot studies. It consisted of eight multi-choice problems, each presenting seven choice alternatives. Each problem outlined a portfolio management scenario, asking subjects to select the alternative that represents their stock investment approach most accurately. Four of the problems dealt with an investment of 100,000 NIS for 1-3 years. The other four problems dealt with an investment of 250,000 for at least 10 years. The two sets of problems were organized in thematic pairs. One of the themes, for example, asked for the number of stock transactions (sale of one stock to purchase another) expected per year, assuming the budget must be directly invested in the local stock market. Figure 3 presents the smaller budget version of the problem.

realized event (hit or miss) in decimal form, was used to derive the winning probability from the likelihood assessments. When drawing the payment assignment, we also drew a 1-100 integer representing the *winning threshold* for the class. The subjects received the 100 NIS when their (payment assignment) winning probability exceeded the threshold.

[9] To decrease the likelihood of zero purchase or sell, the BUY (SELL) assignments dealt with stocks that showed positive (negative) returns at the first half of 2016. The larger (smaller) risk-free rate in the BUY (SELL) problems similarly aimed at decreasing the chances that subjects would choose to buy/sell the 100,000 NIS.

Figure 3: The trading propensity questionnaire – an example

Recall that this chapter deals with the investment of 100,000 in the Israeli stock market for a horizon of 1-3 years ("fixed budget"). This time assume that there are no restrictions on your portfolio management, except for compulsory direct investment in the Israeli stock market. The next alternatives deal with the number of transactions (transaction = purchase of one stock and sale of another) that you would perform in each calendar year. Mark the alternative that suits your investment style best.

1. No more than 1 transaction a year
2. 1-2 transactions a year
3. 3-5 transactions a year
4. 6-8 transactions a year
5. 9-10 transactions a year
6. 11-15 transactions a year
7. At least 15 transactions a year

The three other themes dealt with expected turnover, response to purchase opportunity, and portfolio monitoring style. In analyzing the forecast-confidence and trading propensity correlations, however, we realized that the results slightly strengthen when the portfolio monitoring theme, that can be classified as distinct in scope, is ignored. We therefore measure trading propensity by the other three themes that similarly deal with the tendency to adapt or churn one's portfolio. The pairwise correlations between the six trading items range between 0.19 and 0.63 and Cronbach's alpha is about 0.74. The symbol TP is henceforth used for the normalized 0-100 trading propensity score, with the most extreme subjects at 0 and 100. The main notation is summarized in Table I.

Table I: Notation

FORECAST (F)	The median forecast (abbreviated as F)
R (*r*)	The random (realized) return
HIT	=100 if *r* falls within the $[\mathrm{F}-\delta,\ \mathrm{F}+\delta]$ interval and 0 otherwise
CONF	The 0-100 likelihood assigned to the $[\mathrm{F}-\delta,\ \mathrm{F}+\delta]$ interval
OC	The overconfidence score OC=CONF-HIT
TP	The 0-100 attitudinal trading propensity measure

Notes: The % sign is henceforth suppressed when discussing CONF, HIT and OC

5. Results: forecast-confidence and overconfidence

We briefly report the results of the eight FAATs before proceeding to the confidence-trading correlations. The pairwise correlations between the eight CONFs range between 0.35 and 0.7, with Cronbach's alpha approaching 0.9. The mean CONF is 75, with median 76 and standard deviation 15. Only 7% (N=5) show CONF≤50, while 43% present CONF≥75. Confidence is stronger in the 10% margin tasks compared to the 5% tasks (mean CONFs 78 vs. 72; p<0.01), and also stronger in the NH tasks compared to the H tasks (mean CONFs 77 vs 73; p<0.01), but since the prediction stocks change between tasks we cannot draw conclusions from these comparisons.[10] The mean HIT, however, is only 55, about 20 points smaller than the mean CONF. Almost 90% of the subjects (64 out of 72) show overconfidence in terms of average(CONF)>HIT rate, and the hypothesis that subjects are as likely to overestimate or underestimate their hit rates is easily rejected (p<0.01 by a sign-test). The forecast-overprecision hypothesis is therefore clearly confirmed using the alternative FAAT methodology.[11]

6. Results: the attitudinal confidence-trading link

6.1: The CONF-TP correlations

While we hypothesized that forecast-confidence positively links with trading inclination, the results of two pilot studies surprisingly revealed negative correlations, which reemerged in the current larger experiment. The Spearman correlation between the eight-tasks average CONF and TP is -0.25 and a randomization test (Good, 2013) shows that more negative correlations only arise in 1.8% of 5000 cases when the TP scores are randomly permuted across the sample. A median split of the sample by trading propensity reveals mean CONF of 79 for the subjects with TP smaller than the median, compared to 71 for the others (p=0.02). The correlation gets slightly stronger when confidence is measured by the tasks with no historical charts and it is also stronger for $CONF_{10}$ compared to $CONF_5$ (see Table II). An exhaustive search amongst CONF-related statistics suggested that the minimal CONF in the four tasks with no historical charts (henceforth:

---

[10] Throughout the paper, we use the sign-test for one sample hypotheses and the Pitman test for between samples comparisons. We report Spearman correlations, using randomizations to test significance. Significance levels are 1-tailed and the tables use 3 asterisks *** for p<0.01, ** for p<0.05, and * for marginal p<0.1. We use "average" for within-subject statistics and "mean" for the cross-sample statistics, except for using terms such as "higher than average" in the usual sense.

[11] More details are provided in Sonsino et al. (2020; study 1).

MIN_{NH} CONF) shows the strongest negative correlation, ρ=-0.31 (p<0.01) with TP. The correlation between the minimal CONF in the tasks with historical charts and TP, for comparison, is about half smaller -0.15 (p=0.11).

Table II: The CONF-TP correlations

	CONF	$CONF_{NH}$	$CONF_H$	$CONF_{10}$	$CONF_5$	MIN_{NH} CONF
TP Correlations	-0.25^{**} (p<0.02)	-0.26^{**} (p<0.02)	-0.18^{*} (p=0.06)	-0.26^{***} (p=0.01)	-0.17^{*} (p=0.08)	-0.31^{***} (p<0.01)
Subjects with TP<median (N=36)	79	82	76	82	76	71
Subjects with TP≥ median (N=36)	71	73	69	74	68	57
Pitman test	p=0.01	p=0.01	p=0.03	p=0.01	p=0.03	p<0.01

Notes: The upper panel presents the Spearman correlations between TP and the each of the CONF statistics, using p for the results of a randomization test as described in the text. The lower panel compares the mean confidence of subjects with TP larger and smaller than the median (48.4), disclosing the Pitman test significance levels.

The hypothesis that forecast-confidence associates with stronger inclination to trade at the basic, attitudinal level is therefore clearly rejected. On the contrary, the results suggest that forecast-certainty links with decreased tendency to adapt one's stock portfolio. The negative correlation may seem puzzling at first sight, but becomes plausible at second thought. Types that trust their intuitive (no history) return forecasts consistently, so that even their minimal confidence is relatively high, also exhibit strong confidence in their stock picking, displaying smaller tendency to adapt their holdings. Clearly, such correlation may follow from more basic facets of personality or demographic characteristics that affect trading tendency and forecast-confidence in opposite directions. To search for such underlying effects, we have run separate comprehensive analyses of the factors that affect TP and CONF. The results are summarized next.

6.2: The predictors of trading propensity

The cross-sample analysis pointed at three predictors of individual trading propensity:

6.2.1: Risk-receptiveness

In questionnaire 2, the participants ranked their risk-receptiveness, in personal and professional life, in 1-10 scale. The mean risk-receptiveness (henceforth: RR) was 4.7, with median 4, and rankings smaller than 5 were almost twice more frequent than rankings larger than 5. The discrete

RR emerges as the strongest predictor of trading propensity in the cross-sample analysis. The correlation between RR and TP reaches 0.36 ($p<0.01$), and a median split reveals mean RR of 4.2 for the subjects with lower TP compared to 5.3 for those with higher TP ($p=0.01$).[12] As in diverse preceding studies (Dohmen et al., 2011), the males claim stronger risk receptiveness compared to the females (mean RRs 5 compared to 4.1; $p=0.03$), but the males and females' trading propensities do not differ significantly (mean TPs 49 and 45; $p=0.16$). Risk-receptiveness also strongly correlates with trait optimism as measured in standard LOT-R test ($\rho=0.35$; $p<0.01$), but again optimism does not interact with TP. Links between personal risk-preference and trading inclination were observed in diverse empirical and laboratory studies (Fellner and Maciejovsky 2007; Markiewicz and Weber, 2013; Joao da Gama Batista et al., 2017; Cox et al., 2019). The present results generalize these findings, showing that risk-preference and trading correlate at the attitudinal level.

<u>6.2.2: Perceived competence</u>

Preceding research also indicates that trading volumes may increase with self-confidence and perceived competence (Graham et al., 2009; Bellofatto et al., 2018). Currently, we test for competence effects by asking the participants to Likert rank their perceived abilities in four distinct aspects: (1) academic knowledge in finance (2) familiarity with the local stock market (3) stock selection skills (4) return prediction skills. Indeed, all four measures positively correlate with the trading propensity scores, with the strongest correlations (around 0.28) showing for stated familiarity and stock selection skills. When the familiarity and stock selection scores are averaged to obtain a tentative competence index, the correlation with TP is 0.29 ($p<0.01$) and comparison of the extreme quartiles reveals more than 25% difference in mean trading propensities (mean TPs 41 and 55; $p<0.02$).

[12] Our verbal risk-receptiveness measure is similar to the SOEP measure explored by Dohmen et al., 2011; Crosetto and Filippin, 2016. In addition, Q2 presented a multiple price list task (Andersen et al., 2006) where subjects select between a fixed lottery paying 100 or 200 with equal 50% probabilities and certain payoffs that increase from 100 to 200 in increments of 10. The correlation between the verbal RR and the price list switch point is 0.44 ($p<0.01$), but the switch point shows smaller predictive power for TP (correlation 0.3) and loses significance when RR is controlled.

Table III: The predictors of trading propensity[13]

	RR	Familiarity	C	O	N	LR test
The TP correlations	0.36***	0.29***	0.22**	0.22**	0.11	-
Model (a): T-stats	2.6***	2.1**	2.1**	-	-	21.0***
Model (b): T-stats	2.1**	2.2**	1.9**	0.7		21.5***
Model (c): T-stats	2.8***	1.8**	2.2**	-	1.4**	23.0***

Notes: The first row shows the TP correlations with risk receptiveness (RR), stated market familiarity (Familiarity), Conscientiousness (C), Openness index (O), and Neuroticism (N)**.** Models (a)-(c) summarize the results of ordinal Probit regressions of TP on the standardized variables by disclosing the T statistics of the estimates and the likelihood ratio (LR) of the model. The Probit analysis is selected since TP takes 23 distinct values across the N=72 sample. Similar results emerge in Tobit or OLS regressions, ignoring the discreteness and boundedness of TP.

6.2.3: Personality

We use the 44 problems big five questionnaire developed by John et al. (1991) to characterize the subjects in terms of extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N), and openness (O). Four of the personality traits show positive correlation, exceeding 0.10, with personal trading propensity. The largest correlations, around 0.22, appear for conscientiousness and openness to experience ($p<0.03$). Psychology studies propose that conscientiousness correlates with self-reliance and hard work (Christopher et al., 2008), while openness links to experience seeking and need for cognition (McCrae and Sutin, 2009). The increase in TP with these traits matches these characteristics. The extraversion and neuroticism correlations with TP are also positive (0.16 and 0.11), but statistically insignificant. The agreeableness correlation is close to zero ρ=-0.04. When the personality traits are included as explanatory variables in regressions that account for risk-receptiveness and proclaimed market familiarity, only conscientiousness shows significance (see Model (a) of Table III). Openness to experience strongly correlates with risk-receptiveness (Spearman correlation 0.37) and the openness effect is subsumed by risk-receptiveness when both measures are controlled (see model (b) of Table III). Controlling for risk receptiveness, familiarity and conscientiousness, neuroticism also shows marginally significant positive effect on trading propensity (model (c) of the table). The TP of the subjects that score above average in risk-receptiveness and neuroticism is about 1/3 larger than the TP of the subjects

[13] The regression results are slightly weaker when familiarity is replaced with Competence. We only disclose T-values since the marginal effects are not informative in ordinal Probit regressions.

that score below average in both measures (mean TPs 55 and 42; samples of N=18 and N=23; p<0.01).[14]

6.3: Opposite personality and skills effects on forecast-confidence

In discussing the negative CONF-TP correlations, we suggested that the link may follow from background factors that affect the two variables in opposite directions. Indeed, the cross-sample analysis reveals that openness to experience and neuroticism, that positively interact with trading propensity, negatively correlate with the forecast-confidence statistics. The neuroticism correlations are highly significant, showing for CONF (ρ=-0.38), $CONF_{NH}$ (ρ=-0.29) and MIN_{NH} CONF (ρ=-0.27). The openness correlations are weaker -0.15, -0.20 and -0.22 respectively. The conscientiousness, extraversion and agreeableness correlations are also negative, but smaller, ranging between -0.09 and -0.15. The neuroticism link with confidence clearly reflects in median splits of the sample. The mean CONF of the subjects with smaller than median neuroticism, for example, is about 80 compared to 69 for the others (p<0.01). Comparison of the subjects with below average N and O to those with above average N and O reveals about 1/3 larger MIN_{NH} CONF for the subjects that score low in neuroticism and openness (mean MIN_{NH} CONF 71 compared to 55; samples of N=15 and N=20; p=0.02). Personality studies propose that neuroticism correlates with anxiety (Bishop and Forster, 2013), and links with lower self-efficacy and self-esteem (Judge et al., 2002). The decrease in forecast-confidence with neuroticism fits these general characteristics. The negative openness-confidence correlations intuitively connect to the higher intelligence characterizing open types (Ashton et al., 2000).

In addition, the analysis exposed weak negative correlation between the stated market familiarity or the competence index of section 6.2.2 and the CONF related statistics (see Table IV). The mean MIN_{NH} CONF of the subjects with smaller than median competence, for example, is about 70 compared to 59 for those with larger than median competence (samples of N=31 and N=27;

[14] Positive correlation between openness and risk receptiveness is reported in few preceding studies (e.g., Becker et al., 2012). The evidence on personality effects on financial decision, however, appears inconclusive (see Furnham, 2020 survey) and we could not generally connect the current results to preceding research. Regarding openness, note that Schaefer et al. (2004) find positive correlation between openness and confidence in binary choice general knowledge problems, while we find negative correlation between openness and forecast-confidence. Note also that our Neuroticism correlations with TP and CONF are of the same sign as the E, C, O correlations, although principal component analysis confirmed the existence of a big one trait that negatively loads on N (Musek, 2007).

p<0.05). Again, the effects on confidence are in opposite direction compared to the effects on trading propensity.[15] Models (d) and (e) of Table IV illustrate the significance of the mutual competence and personality effects on forecast-confidence using multivariate regressions.

Table IV: The Predictors of MIN_{NH} CONF

	N	O	Competence	LR-test
The CONF correlations	-0.38^{***}	-0.15	-0.04	-
The $CONF_{NH}$ correlations	-0.29^{***}	-0.20^{**}	-0.07	-
The MIN_{NH} CONF correlations	-0.27^{***}	-0.22^{**}	-0.14	-
Model (d): T-stats	-2.1^{**}		-1.9^{**}	8.9^{**}
Model (e): T-stats	-2.2^{**}	-1.8^{**}	-1.5^{*}	12.0^{***}

Notes: The first 3 rows show the CONF statistics correlations with neuroticism (N), Openness (O), and the competence index of Section 6.2.2. Models (d)-(e) summarize the results of ordinal Probit regressions of MIN_{NH} CONF on the standardized variables by disclosing the T statistics of each estimate and the likelihood ratio (LR) of the model. The Probit analysis is selected since MIN_{NH} CONF takes only 19 distinct values across the sample. Similar results emerge in Tobit or OLS regressions, ignoring the discreteness and boundedness of MIN_{NH} CONF.

6.4: Unraveling the CONF-TP correlation

For Table V, we split the sample between subjects that score relatively high in conscientiousness and openness (the formal criterion is that both normalized scores exceed -0.5) and others. We then median split each sub-sample again by the competence index. Trading propensity increases, while MIN_{NH} CONF decreases, in both horizontal and vertical comparisons, illustrating that the background effects explain the confidence-trading correlation.[16] The correlation between the predicted TP by regression Model (c) (Table III) and the predicted MIN_{NH} CONF by regression Model (e) (Table IV) is -0.61 (p<0.01), about twice larger than the actual correlation (ρ=-0.31). The correlation between the prediction errors of the two models is close to zero ρ=-0.08.

[15] In general, the evidence regarding competence effects on overprecision is mixed. In Glaser et al. (2013), for example, professionals exhibit stronger overprecision than students, while Gloede and Menkhoff (2014) find that overprecision decreases with investment experience.

[16] The split could build on alternative indices; e.g., similar results emerge when the subjects that score high in C, O and N (normalized scores larger than -0.75 in all three traits; N=37) are separated from others.

Table V: The opposite C, N, O and COMPETENCE effects on trading and confidence

	Competence ≤ median	**Competence > median**
High C and O	TP=46 MIN_{NH} CONF = 66 (N=20)	TP=55 MIN_{NH} CONF = 56 (N=20)
Others	TP = 41 MIN_{NH} CONF = 71 (N=15)	TP = 46 MIN_{NH} CONF = 65 (N=17)

Notes: The subjects with normalized C and O scores larger than -0.5 are separated from others and each subsample is split again by Competence. The table shows the mean TP and MIN_{NH} CONF of each group.

7. The case-specific confidence-trading correlations

The negative CONF-TP correlations do not preclude the possibility that confidence-in-profitability motivates trading at the case-specific level. To test the forecast-confidence effect on the stock-specific BUY or SELL decisions, we first use the point forecasts and confidence assessments to estimate the likelihood that subjects assign to profitable purchase or sale. Buy decisions are classified as profitable if the test period return on the stock is positive, while SELL decisions are classified as profitable if the test period return is negative. The likelihood that subjects assign to profitable buy or sell is estimated using normal approximations.[17] If the perceived return is normally distributed around the point prediction F, the probability of return smaller than $(\mathrm{F}+\delta)$ is $(100+\mathrm{CONF})/2$. Assuming that $0<\mathrm{CONF}<100$ for tractability and using Z(p) for the p-th quantile of the standard normal distribution, gives $\delta/\sigma=\mathrm{Z}((100+\mathrm{CONF})/2)$, so that the standard deviation of the perceived return σ can be extracted from CONF and δ. Cases where CONF=100 or CONF=0 are similarly handled assuming CONF=99.99 or CONF=0.01. The estimated σ is then used to calculate the likelihoods of negative or positive return, $\phi\left(\frac{-\mathrm{F}}{\sigma}\right)$ and 1-$\phi\left(\frac{-\mathrm{F}}{\sigma}\right)$, where ϕ represents the cumulative standard normal distribution. Using the symbol CONF+ for confidence in profitability we set CONF+ = $\phi\left(\frac{-\mathrm{F}}{\sigma}\right)$ in SELL problems and CONF+ = $1-\phi\left(\frac{-\mathrm{F}}{\sigma}\right)$ in BUY problems. The top rows of Table VI present the mean CONF+ for each of the four BUY/SELL tasks and the correlation between FORECAST and CONF+ in each problem.

[17] Similar results emerge assuming that the return is uniformly distributed around F. Confidence in profitability could be elicited directly by adding a fourth step to the FAAT, but this could produce spurious correlations between confidence in profitability and the buy/sell amounts.

Since the close to perfect correlations between FORECAST and CONF+ prohibit multivariate regressions, we test the response to CONF+ in two alternative, more direct methods.

Table VI: The CONF+ effects on BUY/SELL amounts

	BUY1	**BUY2**	**SELL1**	**SELL2**
Mean CONF+	69	81	36	57
ρ(F, CONF+)	0.94***	0.86***	-0.93***	-0.99***
Regressions				
FORECAST (F) model	7.1** (4.9)	5.0* (4.0)	-7.1** (3.8)	-14.1*** (6.0)
CONF+ model	11.7*** (5.0)	8.5*** (4.0)	5.3* (3.9)	9.7*** (5.2)
J-test	F rejected for CONF+***	F rejected for CONF+***	CONF+ rejected for F**	CONF+ rejected for F*
Conditional slopes				
Median	1.1*** (p<0.01)	0.65** (p=0.02)	-0.05 (p=0.60)	0.37 (p=0.19)
Weighted average	0.61** (p=0.05)	0.54** (p=0.04)	-0.29 (p=0.88)	0.15 (p=0.38)

Notes: The 4 problems where subjects decided on BUY or SELL amounts are addressed as BUY1, BUY2, SELL1 and SELL2 (see Web supplement 5 for more details). The upmost row of the table presents the mean CONF+ and the correlation between FORECAST and CONF+ in each BUY/SELL assignment. The regressions panel shows the results of separate Tobit regressions (taking into account possible censoring at 0 and 100K) of the BUY or SELL amounts on the standardized FORECAST and the standardized CONF+. The table presents the mean marginal effects with the standard error of the estimate in smaller brackets. The J-test line reports the results of the Davidson and MacKinnon (1981) test for comparing non-nested models. The conditional slopes panel at the bottom of the table shows the median and the weighted average slopes in FORECAST-conditional regressions of the normalized BUY or SELL amounts on the normalized CONF+ measures, as explained in the text and illustrated in Web supplement 7. The smaller brackets p (and asterisks) represent the proportion of cases where larger median or weighted average slopes emerge when the normalized BUY or SELL amounts are randomly permuted within each cluster (samples of 5000 permutations).

First, we separately regress the BUY or SELL amounts on FORECAST and CONF+ and use a test for comparing non-nested models to check if one model can be rejected for the other. The FORECAST and CONF+ variables are standardized for comparability and the J-test of Davidson and MacKinnon (1981) is used to test if the model where the BUY or SELL amounts depend on the standardized forecast can be rejected for the model where the amounts depend on the standardized CONF+ and/or vice versa.[18] The results of the analysis are provided at the regressions

[18] The J-test may bring 4 outcomes: The F model is rejected for the CONF+ model, the CONF+ model is rejected for the F model, neither model is rejected or both models are rejected. See, for example, Bellemare and Barrett (2006).

panel of Table VI. The BUY1 and BUY2 responsiveness to CONF+ (in terms of mean marginal effect) is more than 50% stronger than their responsiveness to FORECAST, and the J-test rejects the FORECAST model for the CONF+ model at $p<0.01$ in both cases. Confidence in profitability thus appears more informative than the point forecasts for the two buy decisions. The SELL tasks regressions, however, show stronger results for FORECAST relatively to CONF+. The FORECAST negative coefficients are larger in absolute value than the CONF+ coefficients in both cases, and the J-test rejects the CONF+ model for the FORECAST model at $p=0.04$ (for SELL1) and at marginal $p=0.10$ (for SELL2).

Alternatively, we take a forecast-conditional approach, clustering the subjects by their point forecasts in each buy or sell assignment, and estimating the responsiveness of the BUY or SELL amounts to CONF+ within each cluster. In BUY1, for example, we separately estimate the slope of the BUY amounts with respect to CONF+ for the two subjects that submitted a point forecast of 15%, the five subjects with FORECAST=10%, the three subjects with FORECAST=8% and so on. The slope calculations are run on individually normalized BUY/SELL and CONF+ statistics, so that the analysis checks if subjects that show strong confidence in profitability, relatively to their personal average CONF+, also choose larger BUY or SELL amounts, relatively to their personal average buy and sell. The number of clusters (with at least two identical forecasts) varied between 12 and 16 across the four tasks, and the total number of observations utilized in calculating the cluster-conditional slopes ranged between 59 and 63 (about 85% of the sample). The lower panel of Table VI presents the median cluster-conditional slopes and the weighted average slopes, where the weighting is based on the number of observations within each cluster (see Web supplement 7 for a detailed example to the calculations). The conditional slope statistics are positive, ranging between 0.54 and 1.1, in the two BUY problems (where slope of 1 implies that 10% increase in the normalized CONF+ increases the normalized BUY by 10,000). The statistics are still positive but smaller (0.37 and 0.15) in SELL2, but SELL1 is an exception with median slope close to zero and a negative weighted average.

To test the significance of the median and weighted average slopes, we use cluster-conditional randomizations, permuting the normalized BUY or SELL amounts within each cluster, independently across clusters. We run 5000 randomizations and count the proportion with median (weighted average) slope exceeding the observed statistics. The results are disclosed in small parentheses at the bottom panel of Table VI. Again, the results are positive for the two buy problems, where less the 1% (2%) of the permuted clustered series showed median slopes larger than the real median in BUY1 (BUY2), and less than 5% of the randomizations displayed weighted average slopes exceeding the real slope in BUY1 and BUY2. While the median and weighted average slopes are still positive for SELL2, the randomizations suggest that even larger median (weighted average) slopes emerge in 19% (38%) of the randomizations, so equality to zero cannot be rejected. For SELL1, the randomizations oppositely suggest that weighted average slopes smaller than observed (-0.29) emerge in only 12% of the randomizations, which comes close to claiming that the weighted average is significantly negative.

Interestingly, both lines of analysis propose that CONF+ is more informative for the amounts that subjects choose to buy compared to the amounts they choose to sell. A possible explanation rests on the observation that CONF+ is about 1/3 smaller in the sell problems compared to the buy problems (means 47 and 75; $p<0.01$). The proportion of CONF+ smaller than 50 is 55% in the sell problems compared to only 9% in the buy problems. The sell amounts of subjects with CONF+ < 50 are significantly smaller than the sell amounts of the subjects with CONF+ ≥ 50 in SELL1 and in SELL2. The low confidence could make CONF+ less relevant than FORECAST for explaining the sell amounts.

8. The conflicting confidence-trading links – discussion

Finance models illustrate that overweighting of private information and underestimation of market volatility may direct traders to excessive non-profitable trading (Daniel and Hirshleifer, 2015). The analysis of the preceding section specifically illustrates that confidence in the profitability of a stock purchase boosts the willingness to buy the stock, beyond the return forecast alone. Forecast-confidence and trading propensity thus positively link at the case-specific behavioral level, when the confidence judgment and trading decision commonly address some stock purchase scenario. More generally, however, the FAAT measure of forecast-confidence links with smaller inclination to churn one's stock portfolio. This part of the results, building on a behavioral forecast-confidence measure on one hand and self-assessed attitudinal trading-propensity score on the other, deals with behavioral and attitudinal dispositions, disconnecting from specific scenarios.

Table VII: Correlations between the behavioral and the case-specific measures

	BUY or SELL	**MIN_{NH} CONF**	**CONF+**
TP	-0.07 (p=0.26)	$\boxed{-0.31^{***}\ (p<0.01)}$	0.13 (p=0.27)
BUY or SELL	-	0.02 (p=0.44)	$\boxed{0.20^{**}\ (p<0.05)}$
MIN_{NH} CONF	-	-	-0.12 (p=0.17)

Notes: BUY or SELL is the average buying/selling amount in the BUY1, BUY2, SELL1, SELL2 tasks. CONF+ is the average confidence in profitability in the respective tasks. TP and MIN_{NH} CONF are as defined in preceding sections. The table present the Spearman correlations for the N=72 subjects; e.g., the -0.07 at the left-top corner represent the correlation between the TP attitudinal measure and the average buy or sell amount in the four specific scenarios.

Table VII displays the correlations between the behavioral and case-specific confidence and trading measures. The bolded boxes present the contrasting behavioral ($\rho=-0.31$; $p<0.01$) and case-specific ($\rho=+0.20$; $p<0.05$) confidence-trading correlations. Interestingly, all the other correlations (relating to the links between behavioral and case-specific measures) do not reach significance. The correlation between TP and the average trading amount in the specific BUY/SELL scenarios, in particular, is close to zero -0.07. Closer analysis, however, revealed that subjects that generally display strong forecast-confidence and accordingly exhibit lower trading propensity, show smaller responsiveness to CONF+ in their case-specific BUY or SELL decisions. To roughly estimate the responsiveness of the BUY/SELL decisions to the likelihood of positive/negative returns, we run subject-level Probit regressions of the BUY/SELL amounts (taking into account possible censoring at 0 and 100,000) on CONF+, building on four observations per subject. We use the responsiveness statistic (the Probit marginal effect; henceforth denoted CONF+_Response) to test if MIN_{NH} CONF correlates with responsiveness to confidence in profitability in the four case-specific decisions. The analysis indeed exposed a weak negative correlation -0.20 (p=0.05) between MIN_{NH} CONF and CONF+_Response, with mean CONF+_Response of 0.16 for the subjects with MIN_{NH} CONF higher than the median compared to 0.48 for those with MIN_{NH} CONF smaller than the median (samples of N=29 and N=26; $p<0.03$).[19] The "stronger forecast-confidence/smaller trading propensity" types still positively respond to CONF+ in their case-level buy or sell

[19] The total BUY/SELL amounts of the two groups do not differ significantly (means 46K vs. 44K; p=0.35) and their average CONF+ are similar (means 59 vs. 63; p=0.14). The mean TPs are 43 and 52 (p=0.02). The significance of the differences in responsiveness is weaker (0.16 vs. 0.52; p=0.12) when the subjects with median MIN_{NH} CONF=70 are included.

decisions, but their responsiveness is 2/3 smaller than the responsiveness of the "weaker forecast-confidence/larger trading propensity" subjects.

9. The results for overprecision

Part of our motivation for introducing FAAT follows from the observation that the task provides a meaningful measure of overconfidence, disentangling the accuracy and confidence effects on forecast-overprecision. The preceding sections of the paper focused on the CONF measures and their links with trading propensity. The current section briefly summarizes the results of a background analysis exploring the accuracy (HIT) and overconfidence (OC=CONF-HIT) correlations with trading. Table IIX presents the simple correlations between (CONF, HIT, OC) and the trade-related variables.

Table IIX: Correlations between (CONF, HIT, OC) and the trading-related measures

	TP	BUY or SELL	BUY or SELL return
CONF	-0.25**	0.03	-0.01
HIT	-0.20*	-0.13	-0.01
OC	-0.06	0.11	-0.01
CONF+	0.13	0.20**	0.11

Notes: The CONF, HIT, OC, CONF+,TP, and BUY or SELL definitions are as in the preceding tables. BUY or SELL return is the average final portfolio return (as defined in 4.3) in the BUY1, BUY2, SELL1, SELL2 tasks.

Our results in terms of the correlation between accuracy or overconfidence on one hand and attitudinal or case-specific trading on the other are not illuminating. The leftmost column of Table IIX shows that the negative -0.25 CONF-TP correlation is offset by negative -0.20 HIT-TP correlation, so that OC=CONF-HIT does not correlate with trading propensity (ρ=-0.06; p=0.3). Closer look at the eight hit rates reveals low pairwise correlations (average phi coefficient 0.028), and Cronbach's alpha of 0.15 compared to 0.90 for the eight CONF assessments. In line with the market efficiency and unpredictability hypotheses (Fama, 2014), the low consistency rates suggest that HIT cannot be adopted as a meaningful forecast-accuracy construct.[20] Given these weak results, we choose not to discuss the negative HIT-TP correlation and the close to zero OC-TP correlation further, just noting again that since forecast-confidence links with smaller tendency to

[20] Predictability is generally stronger for longer horizon and less volatile return targets so that FAAT can generate meaningful forecast-overconfidence measures in other applications.

churn one's stock portfolio our results do not support the hypothesis that types that tend for (exaggerated) forecast-confidence show stronger tendency to adapt their personal stock portfolios. The (CONF, HIT, OC) correlations with the case-specific BUY or SELL amounts or the final return on the BUY/SELL portfolios are also close to zero and clearly insignificant (see the two columns at the right of the table).

With regard to the case specific buying/selling decisions, confidence/overconfidence in profitability may be more relevant than confidence/overconfidence in general. Recall that the CONF+ correlation with the BUY or SELL amount is positive and significant (ρ=0.20; $p<0.05$). The bottom line of Table IIX additionally shows that the CONF+ correlation with the average return on the portfolios that the subjects constructed is also positive (0.11), but insignificant. As CONF+ represents the likelihood that subjects assign to positive or negative return, we accordingly define HIT+ as the proportion of BUY/SELL tasks where the target stock showed positive/negative return, respectively. Examination of the realized returns reveals a HIT+ rate of 50% (as one buy-task stock showed positive return and one sell-task stock showed negative return), compared to mean CONF+ of 60.8%. The hypothesis that subjects (based on the indirect measures) exhibit overconfidence in the profitability of their BUY or SELL decisions is clearly supported (N=57 subjects with CONF+ > 50% and only N=15 with CONF+ < 50%; $p<0.01$). Since, however, the 50% HIT rate applies to all subjects, the correlation between overconfidence in profitability (OC+ = CONF+ minus HIT+) and the trading-related variables is the same as the correlation between CONF+ and these variables (cf., footnote 4). The design of the current survey thus does not allow for comprehensive testing of the hypothesis that overconfidence in profitability links with larger trading and smaller returns.

10. Concluding discussion

The role of personality in economic decision attracts growing interest in recent years (see, Borghans et al., 2008; Becker et al., 2012; Brown and Taylor, 2014; Rustichini et al., 2016; Bucciol and Zaari, 2017; Durand et al., 2019 and Oehler et al., 2020 for diverse examples). The current paper shows that facets of personality that negatively interact with forecast-certainty, positively link with attitudinal trading propensity. Types that show strong belief in the accuracy of their return forecasts exhibit smaller tendency to churn their stock portfolios. Odean (1999), Barber and Odean

(2000) and successive papers illustrate that investors trade stocks too often and returns could significantly improve if portfolios were held for longer periods. Forman and Horton (2019) relatedly show that the round-trip returns of retail Forex traders strongly decrease with the size of their positions. By way of interpretation, the results of the current paper propose that forecast-confident types may suffer less from excessive trading. The results complement a distinct stream in the confidence literature claiming that self-confidence is a prerequisite for success in professional trading and overconfidence may survive in equilibrium (see Oberlechner and Osler, 2012 literature review).

The paper draws a conceptual distinction between the inclination to trade stocks in general and the willingness to buy/sell stocks in specific scenarios. The correlation between our general trading propensity measure (TP) and the case-specific BUY/SELL amounts are close to zero and the smaller tendency of forecast-confident types to churn their portfolios does not reflect in smaller inclination to buy or sell stocks in specific cases. Studies of personality effects on economic decisions argue that situational factors override personal traits frequently, although the predictive power of personality may approach the predictive power of cognitive abilities (Almlund et al., 2011). General traits and situational factors may interact in subtle levels (Mischel, 2004). Our analysis of the conflicting confidence-trading links indeed revealed that the forecast-confident subjects that generally show smaller trading inclination still positively respond to confidence in profitability in their case-specific BUY/SELL decisions, but their responsiveness is 2/3 smaller on average compared to the "low confidence – intensive trading" types.

Part of the innovative approach of this paper is in using interval-evaluation tasks to derive an improved forecast-overconfidence measure. A main advantage of FAAT over standard interval forecasting is in disentangling the accuracy and confidence effects on forecast-overconfidence. Respondents may classify as relatively overconfident for showing smaller than average accuracy, relatively strong forecast-certainty, or both (Sonsino et al., 2020). The analysis in this paper focused on the links between CONF or confidence in profitability and trading-related variables, as the correlations between the HIT or OC scores and the trading variables were not statistically significant. Part of the problem may follow the low internal consistency of the hit rates (Cronbach's alpha 0.15). Another limiting feature is the use of only four case-specific BUY/SELL decisions.

More comprehensive examination of the links between FAAT like-overconfidence measures and trading related variables would require tasks where the hit rates are more meaningful and larger samples of specific buy and sell decisions.

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.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1-181). Elsevier.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383-405.

Ashton, M. C., Lee, K., Vernon, P. A., and Jang, K. L. (2000). Fluid intelligence, crystallized intelligence, and the openness/intellect factor. *Journal of Research in Personality*, *34*(2), 198-207.

Barber, B. M., Lee, Y. T., Liu, Y. J., and Odean, T. (2008). Just how much do individual investors lose by trading?. *The Review of Financial Studies*, *22*(2), 609-632.

Barber, B. M., and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, *55*(2), 773-806.

Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, 4, 453-478

Bellemare, M. F., and Barrett, C. B. (2006). An ordered Tobit model of market participation: Evidence from Kenya and Ethiopia. *American journal of agricultural economics*, *88*(2), 324-337.

Bellofatto, A., D'Hondt, C., and De Winne, R. (2018). Subjective financial literacy and retail investors' behavior. *Journal of Banking & Finance, 92*, 168-181.

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

Bergu, K. (2020). Overconfidence and (Over)Trading: The Effect of Feedback on Trading Behavior. Working paper

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

Bishop, S., and Forster, S. (2013). Trait anxiety, neuroticism and the brain basis of vulnerability to affective disorder. *The Cambridge Handbook of Human Affective Neuroscience*, 553-574.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972-1059.

Broihanne, M. H., Merli, M., and Roger, P. (2014). Overconfidence, risk perception and the risk-taking behavior of finance professionals. *Finance Research Letters*, *11*(2), 64-73.

Brown, S., and Taylor, K. (2014). Household finances and the 'Big Five' personality traits. *Journal of Economic Psychology*, *45*, 197-212.

Bucciol, A., and Zarri, L. (2017). Do personality traits influence investors' portfolios?. *Journal of Behavioral and Experimental Economics*, *68*, 1-12.

Christopher, A. N., Zabel, K. L., and Jones, J. R. (2008). Conscientiousness and work ethic ideology: A facet-level analysis. *Journal of Individual Differences*, *29*(4), 189-198.

Cox, R., Kamolsareeratana, A., and Kouwenberg, R. (2019). Compulsive Gambling in the Financial Markets: Evidence from Two Investor Surveys. *Journal of Banking and Finance, Forthcoming*.

Crosetto, P., and Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, *19*(3), 613-641.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, *1*(2), 115-131.

Daniel, K., and Hirshleifer, D. (2015). Overconfident investors, predictable returns, and excessive trading. *The Journal of Economic Perspectives*, *29*(4), 61-87.

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

Davidson, R., and MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 781-793.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522-550.

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

Durand, R. B., Fung, L., and Limkriangkrai, M. (2019). Myopic Loss Aversion, Personality, and Gender. *Journal of Behavioral Finance*, 1-15.

Fama, E. F. (2014). Two pillars of asset pricing. *American Economic Review*, *104*(6), 1467-85.

Fellner-Röhling, G., and Krügel, S. (2014). Judgmental overconfidence and trading activity. *Journal of Economic Behavior and Organization*, *107*, 827-842.

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

Forman, J., and Horton, J. (2019). Overconfidence, position size, and the link to performance. *Journal of Empirical Finance*, *53*, 291-309.

Furnham, A. (2020). Recruiting, selecting, retaining, and monitoring successful investment managers: A review of a scattered literature. *Financial Planning Review*, *3*(1), e1071.

Gervais, S., and Odean, T. (2001). Learning to be overconfident. *The Review of Financial Studies*, *14*(1), 1-27.

Glaser, M., Langer, T., and Weber, M. (2013). True overconfidence in interval estimates: Evidence based on a new measure of miscalibration. *Journal of Behavioral Decision Making*, *26*(5), 405-417.

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

Gloede, O., and Menkhoff, L. (2014). Financial professionals' overconfidence: Is it experience, function, or attitude? *European Financial Management*, *20*(2), 236-269.

Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359-378.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, *48*(1), 26-34.

Good, P. (2013). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science and Business Media.

Graham, J. R., and Harvey, C. R. (2001). *Expectations of equity risk premia, volatility and asymmetry from a corporate finance perspective* (No. w8678). National Bureau of Economic Research.

Graham, J. R., Harvey, C. R., and Huang, H. (2009). Investor competence, trading frequency, and home bias. *Management Science, 55*(7), 1094-1106

Harrison, G. W., and List, J. A. (2004). Field experiments. *Journal of Economic literature*, *42*(4), 1009-1055.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior and Organization*, *101*, 128-140.

Hey, J. D., and Lee, J. (2005). Do subjects separate (or are they sophisticated)? *Experimental Economics*, *8*(3), 233-265.

Hirshleifer, D., and Luo, G. Y. (2001). On the survival of overconfident traders in a competitive securities market. *Journal of Financial Markets*, *4*(1), 73-84.

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

Joao da Gama Batista, J., Massaro, D., Bouchaud, J. P., Challet, D., and Hommes, C. (2017). Do investors trade too much? A laboratory experiment. *Journal of Economic Behavior and Organization*, *140*, 18-34.

John, O. P., Donahue, E. M., and Kentle, R. I. (1991). The “Big Five” Inventory—Versions 5a and 54. Technical Report. Berkeley: University. Institute of Personality and Social Psychology

Judge, T. A., Erez, A., Bono, J. E., and Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of Personality and Social Psychology, 83*(3), 693-710.

Langnickel, F., and Zeisberger, S. (2016). Do we measure overconfidence? A closer look at the interval production task. *Journal of Economic Behavior and Organization*, *128*, 121-133.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D. (1982). "Calibration of probabilities: The state of the art to 1980," in D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge Univ. Press.

Markiewicz, Ł., and Weber, E. U. (2013). DOSPERT's gambling risk-taking propensity scale predicts excessive stock trading. *Journal of Behavioral Finance, 14*(1), 65-78.

McCrae, R. R., and Sutin, A. R. (2009). Openness to experience. *Handbook of Individual Differences in Social Behavior*, 257-273.

McKenzie, C. R., Liersch, M. J., and Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you?. *Organizational Behavior and Human Decision Processes*, *107*(2), 179-191.

Merkle, C. (2017). Financial overconfidence over time: Foresight, hindsight, and insight of investors. *Journal of Banking and Finance, 84,* 68-87.

Michailova, J., and Schmidt, U. (2016). Overconfidence and bubbles in experimental asset markets. *Journal of Behavioral Finance*, *17*(3), 280-292.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55,* 1-22.

Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502-517.

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.

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, *41*(6), 1213-1233.

Nosic, 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

Oberlechner, T., and Osler, C. (2012). Survival of overconfidence in currency markets. *Journal of Financial and Quantitative Analysis*, *47*(1), 91-113.

Odean, T. (1999). Do investors trade too much?. *The American Economic Review*, *89*(5), 1279-1298.

Oehler, A., Wedlich, F., Wendt, S., and Horn, M. (2020). Does personality drive price bubbles?. Forthcoming, *Studies in Economics and Finance*.

Rustichini, A., DeYoung, C. G., Anderson, J. E., and Burks, S. V. (2016). Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *Journal of Behavioral and Experimental Economics*, *64*, 122-137.

Schaefer, P. S., Williams, C. C., Goodie, A. S., and Campbell, W. K. (2004). Overconfidence and the big five. *Journal of research in Personality*, 38(5), 473-480

Soll, J. B., and Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299-314.

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

Sonsino D., Lahav Y., and Levkowitz A. (2020). "Separating accuracy from forecast-confidence: a modified miscalibration measure." Forthcoming in *Behavioral Finance: a Novel Approach*. Editor, Itzhak Venezia. Publisher, World Scientific.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., and Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis*, *30*(3), 512–23.

Tauni, M. Z., Fang, H. X., and Iqbal, A. (2017). The role of financial advice and word-of-mouth communication on the association between investor personality and stock trading behavior: Evidence from Chinese stock market. *Personality and Individual Differences*, *108*, 55-65.

Teigen, K. H., and Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology*, *19*(4), 455-475.

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.

Winman, A., Hansson, P., and Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1167–1175.

Appendix: The historical charts – an example