Financial professionals' overconfidence: Is it experience, function, or attitude?*

Oliver Gloede †   Lukas Menkhoff ‡§

September, 2011

Abstract

This paper examines financial professionals' overconfidence in their forecasting performance. We compare individuals' self-rating of performance with the true performance, both measured relative to the same peer group. The forecasters in our sample show overconfidence on average, although to a moderate degree, including many cases of underconfidence. In analyzing this, we find that working experience is accompanied by less overconfidence. Function is also related to less overconfidence, such as being a fund manager and using fundamental analysis. The same effect is found for the attitude to herd, whereas recent success appears with more overconfident professionals.

JEL-Classification: G 1, D 84, F 31

Keywords: better-than-average, self-rating, forecasting, foreign exchange, performance

*We thank participants at the annual conferences of the European Financial Management Association, German Economic Association, German Finance Association, Money Macro Finance Research Group, and the Royal Economic Society, as well as the audiences of several seminars, in particular Michael Bräuninger, Jerry Coakley, Michael Frenkel, Markus Glaser, Alan Kirman, Thomas Lux, Christian Pfeifer, Winfried Pohlmeier, Michael Schröder and Philipp Sibbertsen for their helpful comments and discussion. We gratefully acknowledge very constructive comments by the editor and two anonymous referees. In addition, we thank the Centre for European Economic Research (ZEW) in Mannheim, Germany, for providing the data set. Oliver Gloede gratefully acknowledges a scholarship of the Foundation of German Business (Stiftung der Deutschen Wirtschaft)

†Department of Economics, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany, Tel: +49 (511) 762-19960, Email: gloede@gif.uni-hannover.de

‡Department of Economics, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany, Tel: +49 (511) 762-4551, Email: menkhoff@gif.uni-hannover.de

§Centre for European Economic Research (ZEW) in Mannheim, Germany

1 Introduction

We know from a series of, by now, "classical" studies that most people are overconfident regarding their own abilities, such as their driving performance (Svenson, 1981). Subsequent research has shown that overconfidence is not an invariant characteristic but also depends on circumstances. What is important in this respect is the feedback that people receive, which seems helpful in adjusting one's own perception. Financial markets provide frequent and precise feedback to their participants, so that overconfident behavior may be less expected here. In addition, financial markets punish overconfidence, as overconfidence reduces an investor's performance (Odean, 1998). Given the lack of research in this realm, exploring the extent of overconfidence of financial professionals and its potential determinants seems warranted.

Our study contributes to this issue and is the first one to examine determinants of the better-than-average (BTA) variant of overconfidence in the case of financial professionals. Accordingly, we need two ingredients for this study, i.e., a BTA-measure of overconfidence and the linkage of this measure to potential determinants. First, the BTA-measure of overconfidence is the difference between a professional's self-rated relative performance with respect to a reference group and the same person's true relative performance within the same reference group. This precise measure is usually not available. Second, we analyze relationships between individual forecasters' overconfidence and their demographic, (job) function and forecasting characteristics. Findings reported here enrich our understanding about the determinants of overconfidence among financial professionals, perhaps facilitating the introduction of measures for reducing overconfidence and its costly consequences.

Our research uses a sample of 105 professional forecasters who are regularly contributing to an established financial market survey in Germany. From this survey, we gather monthly forecasts for the USD/DM and (later) USD/EUR exchange rate over several years, so that we can calculate a meaningful forecasting performance. We complement this performance information with data from additional surveys conducted at the same time as the regular surveys. These supplementary surveys reveal the professionals' self-rating and further characteristics about them.

The financial professionals in this study are experienced, educated and hold senior positions in the financial industry so that the sample seems to be relevant for our purpose. Based on this sample, the BTA-measure of overconfidence shows reasonable attributes (which will be related to the literature below). First, these professionals regard themselves mostly better in their self-rating than their fellow forecasters and thus show overconfidence on average. Second, the average degree of overconfidence is moderate, which may be expected for professionals who receive frequent feedback and in a market, foreign exchange, where forecasting is particularly difficult. Third, the combination of moderate average overconfidence with substantial individual heterogeneity implies that many professionals show underconfidence. Fourth, the self-rating of financial professionals is not significantly related to the same person's performance, indicating that a correct self-rating is not easy. Fifth, our BTA-measure of overconfidence has the expected correlations to alternative measures of overconfidence and thus seems reliable.

Below we document the following determinants of overconfidence. The working experience of a professional is tentatively related to less overconfidence, suggesting that experience helps in assessing one's true performance. Also, two function-related criteria are related to less overconfidence: first, being a fund manager, which may be due to the immediate feedback that these professionals receive; and second, a heavy reliance on fundamental analysis, which may indicate these professionals' remarkable efforts in forecasting. Finally, a professional's tendency to herd in his forecasts is associated with less overconfidence. Thus, herding may be interpreted as a sign of less confidence and possibly as a rational reaction to inferior information or ability. In contrast to the relations just discussed, recent forecasting success is positively related to overconfidence, with success inducing higher self-evaluation. These patterns provide obvious lessons for financial professionals and their superiors.

This paper is structured as follows: Section 2 reviews the previous literature, showing that our approach is original. Section 3 introduces the data used. Section 4 details relationships between self-rating, overconfidence and performance. Section 5 provides the main findings from regression analyses. And Section 6 concludes.

2 Literature

The aim of this study is to analyze overconfidence of professionals in financial markets. Even though financial professionals are not the main subjects in empirical studies, we can learn about determinants of overconfidence from the finance and psychology literature which deals with overconfidence and BTA in general. A survey of the BTA literature is provided by Alicke and Govorun (2005).

Overconfidence biases are expected to ease when tasks involve frequent feedback, as is the case for financial markets (Lichtenstein and Fischhoff, 1980). The effect of training could be dampened if these biases are deeply rooted in personality (see Preston and Harris, 1965; Brehmer, 1980; Menkhoff and Nikiforow, 2009). Consequently, one should expect a lower degree of overconfidence here than elsewhere. However, the impact of feedback on behavior requires that it is asked for and understood (Kruger and Dunning, 1999). In this respect, there may be differences between various groups in financial markets: professionals in particular might use and be able to interpret the feedback they get. In contrast, non-professional participants, such as individual investors, may have lower incentives to use feedback and may have less ability to interpret it.[1]

Available studies do support the notion that there is also overconfidence in financial markets, but this evidence refers mainly to non-professionals or to financial markets in general, where non-professionals are included. Specifically, there are three kinds of studies. First, one stream of the literature relies on the theoretically-derived finding that overconfidence of financial market participants can be detected by their increased trading activity (Odean, 1998). There is ample evidence of "too much trading volume" in financial markets in general (e.g., Statman et al., 2006). Second, other studies use information about individual investors which is collected from their trading accounts (among others are Odean, 1999; Glaser and Weber, 2007a; Grinblatt and Keloharju, 2009). Third, overconfidence is shown in experiments simulating financial markets (e.g., Biais et al., 2005; Deaves et al., 2009). Thus, we know from these various perspectives that overconfident behavior appears in financial markets, but we know little about whether financial market professionals also show this overconfident behavior.

As professionals trade the largest volumes and have the best information among financial market participants, evidence on their behavior is of particular interest. Empirical examinations, however, lack data. The majority of studies on professionals' overconfidence approach the problem of data availability by developing proxies for overconfidence, such as late option execution (Malmendier and Tate, 2005a), press-related criteria (Malmendier and Tate, 2005b),

[1] In fact, professionals might even need a reasonable level of overconfidence to sustain their optimistic and risk-loving attitude after failure, which is apparently a warranted characteristic of the financial industry (Taylor and Brown, 1988; Oberlechner and Osler, forthcoming).

large numbers of acquisition deals and insider deals (Doukas and Petmezas, 2007), investor size (Ekholm and Pasternack, 2007) and overweighting of private information (Friesen and Weller, 2006). The few studies on financial professionals' overconfidence which employ a direct measure of overconfidence rely on the miscalibration variant of overconfidence (e.g., Ben-David et al., 2007; Deaves et al., 2010) but not on the BTA-measure.

However, the popular miscalibration measure is not without controversy, as different ways of eliciting miscalibration can lead to conflicting results (Cesarini et al., 2006). More important, the various measures of overconfidence are not significantly related to each other. For example, miscalibration is significantly related to neither the BTA-measure (Menkhoff et al., 2006; Glaser and Weber, 2007a) nor the overconfidence measure of illusion of control, whereas BTA and illusion of control are positively related to each other (Menkhoff et al., 2006). Finally, miscalibration is not always related to high trading volume, which is an established theoretical consequence of overconfidence (pro: Deaves et al. (2009), con: Biais et al. (2005); Glaser and Weber (2007a)), whereas BTA is found to have a significant relationship to trading volume (Glaser and Weber, 2007a). In sum, a BTA-measure may provide different information than a miscalibration measure which motivates its application.

There are two studies which are particularly close to ours. First is Deaves et al. (2010), who examine overconfidence among the same group of professionals as we do because both studies rely on the same ZEW data set. In detail, however, there are many differences, such as different time periods, different samples, different financial variables (a stock index vs. foreign exchange rates) and different measures of overconfidence (miscalibration vs. BTA). The second study close to ours is Oberlechner and Osler (forthcoming), who apply a BTA-measure to foreign exchange professionals. Both studies differ in various ways, especially the subject group (foreign exchange traders vs. financial professionals), the research question, and, in particular, the performance measure, as they approximate performance by ratings of superiors and colleagues, whereas we measure "true" forecasting performance by a hit rate.

Overall, studies are rare on financial professionals' overconfidence, in particular regarding the BTA-measure, and existing studies still lack a direct comparison of individual financial professionals' self-rating with the same person's true performance and characteristics.[2]

[2]This approach is not uncommon in the psychology literature though. Several studies have implemented the distinction between 'overestimation' and 'overplacement' (see Moore and Healey, 2008).

3 Data

The study builds on a unique data set which consists of individual exchange rate forecasts over several years plus information about demographic, function and forecasting characteristics of the sample's 105 financial professionals. References to data sources are given in Table A.1 in the appendix.

(Table A.1 about here)

3.1 The ZEW data set

The basis for our research is the individual survey data of the Financial Market Survey conducted by the ZEW in Germany. Overall, there are about 300 financial experts who are asked to participate monthly in the survey, from whom about 250 answers are received each month. Like comparable datasets (e.g., Consensus Economics London), the majority of the participants are employed in the banking sector (75%). Others work in the insurance sector (15%) or in large industrial enterprises (10%). Aggregate statistics of responses are published in financial media like Reuters or Bloomberg. A monthly publication covering the full summary statistics of the survey is also sent to the participating experts, providing aggregate feedback to them.

The surveyed financial professionals provide individual forecasts for the 6-month-ahead exchange rate of the USD/EUR, or of the USD/DM exchange rate before December 1998. The observations range from December 1991 to October 2008. This gives us a maximum of almost 17 years of monthly data on individual expectations. The forecasts are qualitative and indicate whether the exchange rate is expected to appreciate, depreciate or stay unchanged. From these regular forecasts we retrieve average hit rates which approximate the true skill level of the forecasters (for the exact procedure see Section 4). To obtain a reliable and valid criterion for the true skill level, we only use observations from forecasters who participated in the survey at least 36 times (i.e., for a minimum of three years, if they participated every single month).

In addition to this, several special surveys were conducted contemporaneously with the regular monthly surveys, from which demographic and function-related characteristics were obtained (which we describe later in Section 3.3). Since we use personal characteristics and individual forecasts in our analysis, we want to make sure that each observation corresponds to exactly one person. We follow all changes in the contact persons and employers and only use data which refer to the very same person. This provides us with a highly consistent micro data set. The drawback is that we end up with a sample of 105 professional forecasters for whom we have complete observations (i.e., enough forecasts as well as information about their self-rating as forecasters, demographic and job information). Reassuringly, this sample does not show any significantly different items compared to the group who are not considered due to incomplete information.

3.2 Measures of overconfidence

In order to test whether forecasters in the sample exhibit the same behavioral biases as other people, we perform two exercises with conventional results: we observe some overconfidence on average as other studies do, and we find that often-suggested measures of overconfidence are related to each other as in earlier studies. The original survey questions are given in Figures A.1 and A.2 in the appendix.

(Figure A.1 and A.2 about here)

As a first measure of overconfidence, we take the self-rating of respondents and calculate the percentage of financial professionals who think themselves to be better than the average of their peer group. Professionals rank their forecasting performance compared to their participating peers at the ZEW survey on a range from 1 to 21, where "11" represents the average and increasing values represent increasing performance. Earlier studies suggest that there is a tendency to overrate one's own performance. This is usually interpreted as overconfidence of the group on average (Larrick et al., 2007). Recently, this has been called into question. Benoît et al. (2009) show that any fraction could rank themselves as better than average without any overconfident behavior. Our preferred BTA-measure of overconfidence accounts for this by adjusting the self-rating measure of overconfidence for the true performance. Nevertheless, when we employ the widely used aggregated measure of self-rating here, we find the traditional overconfidence pattern, namely that more than 50% of respondents rank themselves as better than or equal to the average forecaster (see Figure 1). Considering the large number of forecasters who give a "being average" rating, the observed level of overconfidence seems moderate (see also Glaser and Weber, 2007b). This may be fostered by three factors. First, the monthly public release of the forecasts gives professionals quite precise feedback about their performance. Second, the more abstract the task is the more overconfident individuals turn out to be (Dunning et al., 1989; Alicke et al., 1995). In the case here, forecasters form concrete expectations about real-world circumstances which may support limited overconfidence. Third, forecasters in our sample do not have an incentive to exaggerate self-rating as the forecasts are anonymous for the public so that the forecasters do not need to fear reputation losses.

(Figure 1 about here)

In order to put overconfidence measures in perspective, we also collect data for the two other measures of overconfidence introduced in the literature section (i.e., miscalibration and illusion-of-control). With regard to the former, in the survey of October 2008, participants were asked for a 90% confidence interval for the 6-month-ahead USD/EUR exchange rate. The mean of the 90% confidence interval stated by the forecasters is about 14 %. This can be compared to two benchmarks: first, we find that the individual confidence intervals are large enough in only 75% of cases compared to realized exchange rates six months ahead. Second, we find that the 14% mean interval width is small compared to the expected interval derived from a GARCH (1,1) model, which is 36%. All this indicates that respondents in our sample tend to be overconfident according to a miscalibration measure. We acknowledge, of course, that this analysis is based on a one-time calibration exercise only. For the third overconfidence measure, illusion-of-control, we collected data asking the following question: "Most of the published business news does not surprise me at all."[3] Respondents answered on a scale ranging from 1 to 20 where "1" gives complete disagreement and "20" gives complete agreement. We find that 80% of respondents answer with categories 11 to 20, thus tentatively supporting the notion that they are not surprised by most news. This provides evidence for overconfidence in the sense of illusion-of-control.

In a final analysis we correlate the three measures of overconfidence introduced above to each other as well as to our BTA-measure (whose exact calculation is introduced later in Section 4).

[3]This question has been used for this purpose before (Menkhoff et al., 2006).

Table 1 shows that these measures are related to each other in a way that is consistent with the literature as discussed in Section 2: in particular, miscalibration measures are uncorrelated with our BTA-measure. All this supports the conclusion that our sample is characterized by similar behavioral biases as found in other studies, although the degree of overconfidence may be relatively small.

(Table 1 about here)

3.3 Measure of performance

We use hit rates as our measure of performance. They are calculated from the raw data as follows. First, we consider all forecasts of one person. Second, we determine exactly whether a particular forecast was right or wrong. Survey participants have a time window of about two weeks to submit their forecasts. To achieve a maximum of accuracy and consistency we use individual-specific forecasting days. Specifically, we compare the forecasted change of the exchange rate to the realized exchange rate in exactly six months for each individual separately. Third, since the expectations are qualitative forecasts, usual error measures (e.g., RMSE) are not computable, necessitating the use of hit rates. For this purpose we convert the continuous exchange rate process into a discrete process which corresponds to the forecast categories of appreciation, depreciation and no change. We use information directly from the forecasters themselves. In a special survey in 2006, they state that, on average, a plus or minus 3% change of the exchange rate over six months is considered to be stable. Fourth, to incorporate the fact that the experts can choose between three alternatives, a hit rate is coded in three categories: a large deviation, a small deviation and no deviation of forecast from the true event. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. Code values of 0-1-2 are utilized where a higher hit rate implies greater success.[4]

(Table A.2 about here)

[4]This performance measure has been used previously for ZEW exchange rate forecasts (e.g., Nolte et al., 2008). Details about the calculation of the hit rates are given in Table A.2 in the appendix.

3.4 Professionals' characteristics

In order to examine potential determinants of overconfidence, the ZEW individual forecasting data have to be supplemented with detailed information about the financial professionals. We do this in three directions, addressing professionals' demographic, function-related, and forecasting characteristics.

(Table 2 about here)

Regarding demographic characteristics, the average professional is male, has an academic education, is middle-aged and has been working for almost 18 years in the financial industry. Descriptive statistics for the sample of 105 financial professionals are shown in the upper panel of Table 2. Statistics are consistent with the information provided by Deaves et al. (2010) based on the year 2003.

Regarding function characteristics the average professional holds a senior position, exercising operative and personnel responsibilities. It is most likely that he works as a fund manager (30%); another 23% work as researchers; 18% as advisors; and the remaining 29% are classified as others.[5] Forecasts are mainly the result of fundamental analysis, which has a share of 55%, whereas technical analysis and reliance on order flow analysis make up the rest (the survey question is documented in Figure A.1). Detailed results are given in the middle panel of Table 2.

Regarding forecasting-related characteristics, we first report the average hit rate, which can vary between a lower bound of zero and an upper bound of two as discussed in Section 3.3 above. Beyond that we are interested in two possible behavioral effects: first, do recent forecasting successes lead to increased self-assessment? Second, is the financial professional inclined to herd in his forecasts?

We investigate the impact of recent forecasting success with two variables. First, in order to obtain a measure of the forecasters' recent success, we measure whether there is a significant - positive or negative - trend in the performance over the last three years, prior to asking for the respondent's self-rating. To calculate this trend over three years we use a simple Spearman

[5] Such as employees in treasury departments.

rank correlation between the numbers 1 to 36 and the hit rates at these 36 points in time. The coefficient of correlation can be positive or negative, but only if it is significant at a 10% level do we take it as a trend. Then we form two dummy variables: the dummy is one for a positive (negative) trend (i.e., increasing (decreasing) success in forecasting), and zero for all other cases. Table 2 shows that 15% of respondents experienced a positive trend, 7% a negative trend and most respondents did not realize any trend.[6] Second, as another proxy of recent forecasting success, we calculate the average individual hit rates during the six months prior to the self-rating. In order to get six observations per person, we compare the original 6-month forecasts with the actual movement in the exchange rate within the first month after the forecast was made. For this procedure we adjust the no-change category according to the square root formula (i.e., the 6-month boundary of 3% corresponds to a 1-month boundary of 1.22%). Our measure of recent success is the average of these newly computed recent hit rates of the last 6 months, which is the time horizon shown to be relevant by Statman et al. (2006) and others. Reassuringly, our results are robust for other time periods such as 5 or 7 months. The level of recent success as shown in the lower panel of Table 2 is not significantly different from the general performance.

The last variable in the lower panel of Table 2 provides information on potential herding behavior among forecasters. We implement as a herding measure the simple percentage share of one's monthly forecast in which the professional agrees with the market's opinion of the month before. The market opinion is approximated here by the mode of responses. To ensure a robust estimate of the market opinion, we choose the minimum participation rate to be 30, which can be important during the very early years of the survey.[7] The bottom row of the lower panel of Table 2 shows that about half of the time professionals conform to the market opinion.

4 Descriptive analysis

To review, our BTA-measure of overconfidence is defined to be the difference between self-rating and performance. We show the distribution for these two ingredients of overconfidence in the

[6] Alternatively, we replace the Spearman rank correlation coefficient by Kendall's tau but obtain the same signs. Results also remain qualitatively the same for variations in the time horizon.

[7] Results are basically unchanged if we either pick another minimum participation rate or if we take the contemporaneous month.

sample and how they are related to overconfidence.

In Figure 2 we plot the 105 forecasters' self-rating against their true performance (measured via hit rates). A relation between self-rating and performance cannot be recognized, either from graphical analysis or from the correlation coefficient (-0.0084). This result may seem a bit surprising because financial professionals receive frequent feedback about their performance and because their salary is performance-based to some degree. Consequently, one might expect that they are tentatively able to correctly self-assess their performance. However, the realized hit rates suggest that professionals cannot really forecast exchange rates very well. They may be more successful at longer horizons (Heiden et al., 2011), but they do not succeed on average at the medium-term horizons used here (e.g., Ruelke et al., 2010). In sum, it seems difficult also for the individual forecaster to develop a precise relative self-assessment.

(Figure 2 about here)

Next, we relate our BTA-measure of overconfidence to its two ingredients (as described above) and see which ingredient may be driving overconfidence.[8]. Thus in order to calculate the BTA-measure, its two ingredients should be made comparable to each other regarding their scaling. Therefore, hit rate is linearly adjusted to the same scale as self-rating. Hit rate is initially a continuous variable ranging from 0 to 2. For adjustment to the self-rating scale (from 1 to 21), we take the mean, which is 1.13, and take four standard deviations to both sides to decide the boundaries: the lower bound is 0.548 and the upper bound is 1.727. These boundaries fully encompass the actual hit rates, which range from 0.577 to 1.510. Then, we split the defined range into equal parts from 1 to 21 so that the hit rate is now easily comparable to the self-rating. The resulting BTA-measure of overconfidence (self-rating minus transformed hit rate) has observed values between -13 and 14, which means that on an individual basis professionals can exhibit either overconfidence or underconfidence.

(Figure 3 about here)

[8] Due to the labeling of the figure in the survey question we can relate the best and worst categories of the hit rate and self-rating as well as the average to the middle category. We cannot be sure that the self-rating categories in between match the corresponding hit rate categories. The empirical approach tries to accommodate for that by estimating just three different categories instead of estimating more categories.

Figure 3 provides a plot of the relationship between overconfidence and the hit rate for the sample of 105 forecasters. Obviously, a better hit rate, i.e. a better forecasting performance, goes along with less overconfidence. Additionally, we provide the analogous plot for overconfidence versus self-rating. Figure 4 shows that forecasters are often overconfident when giving relatively positive self-assessments. In contrast, relatively underconfident forecasters tend to give pessimistic self-ratings.

(Figure 4 about here)

5 Regression analysis

5.1 Methodology

We next seek to explain the overconfidence (or underconfidence) of the professional forecasters in our sample by a set of demographic, function-related, and forecasting characteristics. An ordinal logit model is estimated where over-/underconfidence is defined as a piecewise-defined function of the difference between self-rating and hit rate as follows:

$$OVC_i = \begin{cases} 1 & \text{if } SR_i - HR_i > 0 \\ 0 & \text{if } SR_i - HR_i = 0 \\ -1 & \text{if } SR_i - HR_i < 0 \end{cases} \tag{1}$$

with SR_i as self-rating and HR_i representing hit-rate, each for forecaster i. Thus, the criterion distinguishes overconfident and underconfident forecasters as well as forecasters who are neither. We observe 51% overconfident, 36% underconfident and 12% "balanced" professionals.

Using this non-linear estimation procedure reflects the nature of our data and is thus more appropriate than an ordinary least squares analysis. First, using the directional information of the difference between self-rating and transformed hit rates puts less restrictive assumptions on the data than using a cardinal criterion. In a strict sense, the self-rating scale is only well-defined at the average and at the endpoints, whereas the information about ratings between these points cannot be directly compared across persons, as an OLS analysis assumes. Second, the ordinal logit allows for the possibility that the effect on over- and underconfidence of the regressors could be asymmetric, i.e. nonlinear. To control for heteroscedasticity we use robust variance estimators for all estimations.

5.2 Results

We estimate the level of over- / underconfidence conditional on a large number of control variables which we derive from demographic, function, and forecasting characteristics. We find that experience, function-related, and forecasting attitudes are significantly related to the level of overconfidence.

Estimation results are shown in Table 3. Column (1) provides the benchmark specification where all demographic, function-related and forecasting characteristics are considered. We take the results of the restricted regression in column (2) for interpretation which only uses recursively significant regressors. Additionally Table 4 shows marginal effects. We discuss explanatory variables by category, starting with the demographic characteristics. Note that specifications (3) and (4) in Table 3 reproduce specifications (1) and (2) with the only difference being the hit rate trend variables (in (1) and (2)) are replaced by a single variable indicating (six-month) recent success.

(Table 3 and Table 4 about here)

(1) Demographic characteristics. Bhandari and Deaves (2006) show that demographic characteristics can indeed influence financial professionals' behavior. Among the demographic variables available for our sample, we find that experience has a significant effect, which is negatively related to overconfidence. The direction of this effect is not as obvious as it may look at the first sight. During one's career, a forecaster achieves both successes and failures. Using this information allows her to form a rational expectation about her own forecasting performance even when the true skill was unknown in the beginning. This positive learning effect of experience may be tempered by forces preventing forecasters from learning, such as confirmatory bias and self-attribution bias (Brehmer, 1980). These forces conceivably may swamp learning, leading to greater overconfidence for more experienced forecasters. These positive and negative influences of experience on overconfidence have been formalized in a multi-period model by Gervais and

Odean (2001), who argue that, typically, experience supports the development of overconfidence in the early stages of one's career but then later on it will depend on the level of self-attribution bias as to whether more experience will lead to either lower or higher overconfidence. Thus, theory states that the relation between experience and overconfidence may differ between persons and may even change over time for the same person.

The evidence on an experience effect is indeed mixed. For example, Oberlechner and Osler (forthcoming) do not find a significant learning effect. Providing evidence in favor of the learning effect, Glaser and Weber (2007b) find that experienced private investors are better able to self-evaluate their portfolio returns than inexperienced investors. Our evidence shows that working experience in the financial sector is associated with reduced overconfidence at the 5% significance level. This result is independent of controlling for age, so the experience effect does not stem from just getting older.[9] In Table 4 we see that for an average forecaster eight more years (one standard deviation) of experience lead to a 12% lower probability of being overconfident and to a 10% higher probability of being underconfident.

Our finding on experience confronts the result of Deaves et al. (2010) in whose study experience increases overconfidence. As mentioned in the literature section both studies use the same survey, however, studies differ regarding the time period covered, the sample definitions, the financial markets covered (stocks vs. foreign exchange) and the measures of overconfidence (miscalibration vs. BTA). Because of these various differences between both studies one cannot identify a single (most important) reason for the contrary relation of experience to overconfidence. A plausible reason might to be, however, that the measures of overconfidence themselves are not positively correlated to each other so why should experience be related to them in the same way? Possibly BTA biases can be eased by experience when miscalibration cannot.

Since gender is a frequently discussed issue in the related literature (e.g., Barber and Odean, 2001), we control for gender effects. We cannot find any significant difference between the behavior of women and men. Due to the large fraction of men (92%) in our sample we do not draw any conclusion from this result.

(2) Function-related characteristics. Besides the information about demographic attributes we analyze the influence of function-related characteristics by including dummies for

[9] The effect remains if we use othogonalized variables (available on request).

advisor, researcher and fund manager in the benchmark regression. Fund managers are more (less) likely to be underconfident (overconfident) by 32% than non-fund managers. This may be due to the direct feedback which fund managers receive. Among our respondents, they are the financial market participants with the clearest direct feedback and their salary is usually linked to their performance.

Besides the position dummy for fund managers, we find a significantly negative effect for the heavy use of fundamental analysis. We interpret this variable as a measure of the extent to which one uses complex analytical methods rather than simple technical rules or relying just on good luck. Former research has shown that sophistication can decrease biases (Feng and Seasholes, 2005). Nevertheless, the effect of this variable is not robust for all regressions (see specification 4 in Table 3).

(3) Forecasting characteristics. In the theoretical literature, overconfidence is modeled as a process of learning due to biased self-attribution (e.g. Daniel et al., 1998; Gervais and Odean, 2001): recent successes take relatively too much weight for self-evaluation (Miller and Ross, 1975). We observe both positive as well as negative trends in the forecasting performance of the last three years. If self-attribution bias is a reason for overconfidence, a positive trend in performance should be significantly related to overconfidence, while a negative trend should have no impact. Indeed, this is what we find. The dummy variable for a positive trend in the hit rate is significant for all model specifications. The analysis for the reference case of an average forecaster shows that recent success measured in this way results in a 27% higher chance of being overconfident and reduces the likelihood of being underconfident by 19%.

In narrowing the time frame for possibly biased self-attribution, one can test whether the most recent successes are also important for understanding overconfidence. Using aggregate data Statman et al. (2006) find that returns going back 6 months matter for trading volume, which they interpret as overconfidence. We also provide support for this relationship. We estimate an adjusted model which includes the mean hit rate of the last 6 months rather than the dummy variables for positive and negative trends in the hit rate (see specifications 3 and 4 in Table 3). The coefficient for the new variable of recent success is significantly positive, which supports the role of self-attribution bias. Recent successes seem to generate an attitude which biases forecasters' self-evaluation positively.

Turning to the tendency to herd, we measure how much each forecaster agrees in her forecasts with the market opinion (the mode of forecasts in our sample) and interpret this as herding behavior. We observe that the more (less) the forecaster aligns his forecast with the market opinion, the less (more) overconfident (underconfident) this professional is. This could be a rational reaction to inferior information or ability. Another explanation for this relationship may focus on forecasters' risk attitude which influences both herding and overconfidence. A herding forecaster tends to rate herself quite conservatively due to high risk aversion. Theoretical studies show that, due to reputation effects, lower risk taking and more intensive herding go hand-in-hand (Hirshleifer and Thakor, 1992). Empirical evidence for this relationship is abundant, including for example Graham (1999). Therefore the relationship between herding and overconfidence may stem from the fact that they are both influenced by risk aversion. An alternative explanation for the observed link between herding and overconfidence could be driven by the performance of the market opinion. If the market opinion is a better forecast than the average of the individual forecasts, herders will show up as comparatively good forecasters. This leads ceteris paribus to our observation that herders are less overconfident. Testing this hypothesis, we calculate the hit rate of the market opinion, which is 1.39 over the whole time span. Comparing that to the average forecasting performance of 1.13, we indeed find that the market opinion is significantly better than the average hit rate of the individual forecasts. This finding supports the alternative argument that a relatively precise market opinion explains the link between herding and less overconfidence.

5.3 Robustness

We next report several robustness tests relating to sample selection, alternative measurement of hit rates, different threshold levels (for perceived unchanged exchange rates), various further regression models, and, finally, different transformations of hit rates.

(1) Sample selection. As we are restricted to working with a sample of 105 financial professionals out of a total of more than 300 respondents to the monthly survey, the issue of representativeness must be addressed. The main reason that the sample is so much smaller than the number of respondents is the unavoidable reliance on questionnaire responses, additional to the regular survey. These additional questionnaires are necessary to obtain information, first, about self-rating, and, second, about various demographic and function-related characteristics. A third restriction results from our requirements that only persons with at least 36 months of observations are included and more than three years of observations are necessary in order to calculate individual trends in hit rates. Although we did manage to obtain more than 200 responses from the additional questionnaires, the combination of requirements reduces the sample to 105 professionals.

In order to test unbiasedness of this sample, we compare means between the 105 professionals included and those professionals who had to be excluded. Table 5 provides the results. Panel A reports comparisons for all those variables described in Table 1 above and Panel B report comparisons for other variables of interest. Importantly, there is no single significant difference in variables' means between our sample and the group of excluded professionals.[10]

(Table 5 about here)

Sample selection bias could also exist due to panel attrition. This is because forecasters drop out of the sample occasionally and are replaced. To analyze the effect of the duration of panel affiliation we correlate our overconfidence measures to two measures which indicate how long a financial expert has belonged to the panel.

(Table 6 about here)

Neither the duration of how long an expert participated in the survey nor the number of forecasts the person gave during this time are correlated with the overconfidence measures. We conclude that our results are unlikely to be biased by panel attrition.

(2) Hit rate calculation. For our baseline estimations we use a three-variate hit rate with three codes: success, small failure, and large failure. Since we also receive the forecasts in three different outcomes, this procedure seems quite reasonable. For robustness we estimate our baseline model again, using a hit rate which distinguishes only between giving a correct

[10]It would be preferable to apply a formal Heckman model to test for sample selection and correction, if necessary. However, due to the incomplete data set, there is no group of excluded professionals where we would have sufficient information to identify the model for the full sample. Trying various nested-models identification improves but the Mill's Ratio is not significant in any case. This suggests that we do not have a serious problem of sample selection.

and an incorrect forecast (see Table 7). Thereby, we obviously discriminate forecasters who make just small mistakes, but do not get the direction wrong. Despite this mistake, our results remain generally the same. Forecasters still do not get their self-rating right, which means that self-rating and performance are uncorrelated. On average, forecasters are truly overconfident. The determinants of overconfidence also remain the same. Working experience as well as the forecasting characteristics remain significant and keep their effect. The effect of being a fund manager also remains stable, whereas the effect of fundamental analysis vanishes. This underlines the low significance level of fundamental analysis in our baseline model and encourages us not to overestimate the effect of fundamental analysis.

(Table 7 about here)

(3) Different threshold levels. A possible shortcoming of the above analysis is its reliance on a fixed average threshold of plus or minus 3% for forecasters' perception of unchanged exchange rates. The 3%-level is chosen because it represents the mean (and median) derived from participants in the ZEW Financial Market Survey.[11] In order to address possible variation over individuals and time, we recalculate the main regressions with average levels of 2% and 4%, which seems appropriate since 70% of individual responses fall between these percentages. The new regressions shown in Tables 8 and 9 show that our findings are quite robust to these variations.

(Table 8 and 9 about here)

(4) Further regression models. Next, we implement further regression models. First, we relax our assumption of a logistic distribution of the error terms. The information criteria suggest that the logistic assumption fits our data set quite well, as the Akaike Information Criterion and Bayes Information Criterion are both fairly smaller for the ordered logit model than for an ordered probit model. Nevertheless, when we use the normal error distribution assumption, the significance levels of the parameter estimates remain the same (not shown to save space).

[11] As individual survey responses are available, the individual threshold levels could be used in principle. However, the benefit in precision is limited by reduced sample size (48 persons only) and possibly time-varying threshold values which we cannot account for.

Second, we test for the proportional odds (also parallel regression) assumption of the ordered logit model and use alternative models which relax this assumption. A likelihood ratio test with the null hypothesis of proportional odds shows no evidence for a violation of the assumption for our data set. Testing the assumption of proportional odds for each coefficient individually, we find that a few variables violate the assumption. Therefore we relax the parallel regression assumption and estimate a partially generalized ordered logit model (see Table 10). This method allows us to lift the constraint for some variables and to restrict the rest. For most of the variables, the results remain the same. But the estimation suffers from a large proportion of negative-predicted probabilities, which accounts for about two-thirds for some specifications. Moreover, the estimates are blurred due to the opaque impact of some variables, an example being gender with only 7 % of the sample being female. Hence, we prefer the ordered logit model compared to the generalized ordered logit model.

(Table 10 about here)

Third, as a further robustness test, we introduce a model which removes the assumptions of the ordinal features of the data and estimate a multinomial regression model. For this estimation approach, the results also remain mainly the same. Since the multinomial estimation clearly violates the results of the likelihood ratio test and neglects the ordinal nature of the dependent variable, the ordered logit model is our preferred model. As an even more radical departure from our preferred estimation approach, we neglect the ordinal character of our data running an OLS regression for comparison purposes. We find that coefficient signs remain, supporting the robustness of findings, but that significance levels go down or even disappear (available on request).

(5) Different hit rate transformations. Finally, we acknowledge that the hit rate transformation we used in Section 4 may be questioned. As a first alternative, we replace our transformation of hit rates into 21 equal parts into a sorting of respondents into 21 quantiles (as suggested by a referee). We prefer the linear transformation as this maintains a normal distribution for hit rates, which also fits the approximately normal distribution of self-ratings (see Figure 2). By comparison, the transformation of hit rates into quantiles implies equal use of the full scale (i.e., a distribution which is in principle possible but not really suppported by our data). Results are shown in Table 11 and indicate robustness of the overall findings. However, some variables become less significant (experience and fundamental analysis) or lose significance (recent success).

(Table 11 about here)

In further exercises, we maintain the linear transformation of hit rates but apply different band widths. For example, when we replace the four-standard deviation band by a three-standard deviation band (in order to reduce the impact from extreme values) or by a five-standard deviation band (in order to leave room for a more extreme outcome not represented in our limited sample), we find that coefficient signs remain the same but the levels of significance go down (available on request). However, this is to be expected, as the estimation does not optimally use the variance in observations.

6 Conclusion

This study examines overconfidence (and underconfidence) among financial professionals. We contribute to the literature in that we combine "hard" performance information with self-rated performance and complement this with a comprehensive set of demographic, function-related and forecasting characteristics. Further, the utilization of a BTA-measure of overconfidence measure among financial professionals strikes new ground.

We find that financial professionals in our sample are overconfident on average, although the degree of overconfidence seems relatively small compared to many studies with individual investors. Moreover, we find that the positive relationship between self-rating and performance is not statistically significant, which may be a bit surprising for professionals. Consequently, overconfidence is driven by high self-ratings and low performance. Interestingly, there are also underconfident professionals who have been largely neglected in earlier research.

In an effort to understand financial professionals' over- and underrating of their own performance, we examine correlates suggested in the literature. We find that experience in the financial sector is associated with less overconfidence. Some function-related variables, such as being a fund manager and the use of fundamental analysis, are also related to less overconfidence. Finally, recent success and non-herding are observed among more overconfident financial professionals.

Overall, these intuitively plausible results contribute to our understanding of overconfidence among professionals. They also indicate ways to limit its adverse consequences: for example, reliance on more experienced professionals could be helpful in this respect, as well as giving frequent and precise feedback about performance. In addition, debiasing training may be called for. A final contribution in limiting overconfidence may lie in clearly distinguishing between forecasting performance and marketing performance, as the latter needs bold forecasts to create attraction, whereas the former may profit from moderate forecasts.

References

Alicke, M. and O. Govorun, "The Better-Than-Average Effect," in M. Alicke, D. Dunning, and J. Krueger, eds., *The Self in Social Judgement*, Studies in Self Identity, Psychology Press, 2005, pp. 83–106.

Alicke, M.D., M.L. Klotz, D.L. Breitenbecher, T.J. Yurak, and D.S. Vredenburg, "Personal Contact, Individuation, and the Better-Than-Average Effect," *Journal of Personality and Social Psychology*, 1995, *68*, 804–825.

Barber, B.M. and T. Odean, "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 2001, *116*, 261–292.

Ben-David, I., J.R. Graham, and C.R. Harvey, "Managerial Overconfidence and Corporate Policies," *NBER Working Paper*, 2007, *13711.*

Benoît, J.-P., J. Dubra, and D. Moore, "Does the Better-Than-Average Effect Show That People Are Overconfident? An Experiment," *MPRA Working Paper*, 2009, *13168.*

Bhandari, G. and R. Deaves, "The Demographics of Overconfidence," *Journal of Behavioral Finance*, 2006, *7*, 5–11.

Biais, B., D. Hilton, K. Mazurier, and S. Pouget, "Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market," *Review of Economic Studies*, 2005, *72*, 287–312.

Brehmer, B., "In One Word: Not from Experience," *Acta Psychologica*, 1980, *45*, 223–241.

Cesarini, D., Ö. Sandewall, and M. Johannesson, "Confidence Interval Estimation Tasks and the Economics of Overconfidence," *Journal of Economic Behavior and Organization*, 2006, *61*, 453–470.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, "Investor Psychology and Security Market under- and Overreactions," *Journal of Finance*, 1998, *53*, 1839–1885.

Deaves, R., E. Lüders, and G.Y. Luo, "An Experimental Test of the Impact of Overconfidence and Gender on Trading Activity," *Review of Finance*, 2009, *13*, 555–575.

— , — , and M. Schröder, "The Dynamics of Overconfidence: Evidence from Stock Market Forecasters," *Journal of Economic Behavior and Organization*, 2010, *75*, 402–412.

Doukas, J.A. and D. Petmezas, "Acquisitions, Overconfident Managers and Self-Attribution Bias," *European Financial Management*, 2007, *13*, 531–577.

Dunning, D., J.A. Meyerowitz, and A.D. Holzberg, "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability," *Journal of Personality and Social Psychology*, 1989, *57*, 1082–1090.

Ekholm, A. and D. Pasternack, "Overconfidence and Investor Size," *European Financial Management*, 2007, *14*, 82–98.

Feng, L. and M.S. Seasholes, "Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets?," *Review of Finance*, 2005, *9*, 305–351.

Friesen, G. and P.A. Weller, "Quantifying Cognitive Biases in Analyst Earnings Forecasts," *Journal of Financial Markets*, 2006, *9*, 333–365.

Gervais, S. and T. Odean, "Learning to Be Overconfident," *Review of Financial Studies*, 2001, *14*, 1–27.

Glaser, M. and M. Weber, "Overconfidence and Trading Volume," *Geneva Risk and Insurance Review*, 2007, *32*, 1–36.

— and — , "Why Inexperienced Investors Do Not Learn: They Do Not Know Their Past Portfolio Performance," *Finance Research Letters*, 2007, *4*, 203–216.

Graham, J.R., "Herding among Investment Newsletters: Theory and Evidence," *Journal of Finance*, 1999, *54*, 237–268.

Grinblatt, M. and M. Keloharju, "Sensation Seeking, Overconfidence, and Trading Activity," *Journal of Finance*, 2009, *62*, 549–578.

Heiden, S., C. Klein, and B. Zwergel, "Beyond Fundamentals: Investor Sentiment and Exchange Rate Forecasting," *European Financial Management*, 2011.

Hirshleifer, D. and A.V. Thakor, "Managerial Conservatism, Project Choice, and Debt," *Review of Financial Studies*, 1992, *5*, 437–470.

Kruger, J. and D. Dunning, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 1999, *77*, 1121–1134.

Larrick, R.P., K.A. Burson, and J.B Soll, "Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence (and when It Does not)," *Organizational Behavior and Human Decision Processes*, 2007, *102*, 76–94.

Lichtenstein, S. and B. Fischhoff, "Training for Calibration," *Organizational Behavior and Human Performance*, 1980, *26*, 149–171.

Malmendier, U. and G. Tate, "CEO Overconfidence and Corporate Investment," *Journal of Finance*, 2005a, *60*, 2661–2700.

_ and _ , "Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited," *European Financial Management*, 2005b, *11*, 649–659.

Menkhoff, L. and M. Nikiforow, "Professionals' Endorsement of Behavioral Finance: Does it Impact their Perception of Markets and themselves?," *Journal of Economic Behavior and Organization*, 2009, *71*, 318–329.

_ , U. Schmidt, and T. Brozynski, "The Impact of Experience on Risk Taking, Overconfidence and Herding of Fund Managers: Complementary Survey Evidence," *European Economic Review*, 2006, *50*, 1753–1766.

Miller, D.T. and M. Ross, "Self-Serving Bias in Attribution of Causality: Factor Fiction?," *Psychological Bulletin*, 1975, *82*, 213–225.

Moore, D.A. and P.J. Healey, "The Trouble with Overconfidence," *Psychological Review*, 2008, *115*, 502–517.

Nolte, I., S. Nolte, and W. Pohlmeier, "The Good, the Bad and the Ugly: Analyzing Forecasting Behavior within a Misclassified Quantal Response Framework," *University of Konstanz Working Paper*, 2008.

Oberlechner, T. and C. Osler, "Survival of Overconfidence in Currency Market," *Journal of Financial and Quantitative Analysis*, forthcoming.

Odean, T., "Volume, Volatility, and Profit When All Traders Are above Average," *Journal of Finance*, 1998, *53*, 1887–1934.

_ , "Do Investors Trade Too Much?," *American Economic Review*, 1999, *89*, 1279–1298.

Preston, C.E. and S. Harris, "Psychology of Drivers in Traffic Accidents," *Journal of Applied Psychology*, 1965, *49*, 284–288.

Ruelke, J.C., M.R. Frenkel, and G. Stadtmann, "Expectations on the yen/dollar exchange rate - Evidence from the Wall Street Journal forecast poll," *Journal of the Japanese and International Economies*, 2010, *24* (3), 355 – 368.

Statman, M., S. Thorley, and K. Vorkink, "Investor Overconfidence and Trading Volume," *Review of Financial Studies*, 2006, *19*, 1531–1565.

Svenson, O., "Are We All Less Risky and More Skillful Than Our Fellow Drivers?," *Acta Psychologica*, 1981, *47*, 143–148.

Taylor, S.E. and J.D. Brown, "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, 1988, *103*, 193–210.

Figure 1: Histogram of financial professionals' self-rating

This figure presents the distribution of financial professionals' self-rating. Self-rating is the individual response on a survey question, which we asked the financial experts in two surveys (April 2007 and October 2008): "How do you evaluate your USD/EUR-forecast compared to the average forecasting hit rate of all participants of the ZEW financial market survey?". The scale ranges from 1 to 21. The mean answer is 11.9, which is above the theoretical mean of 11.

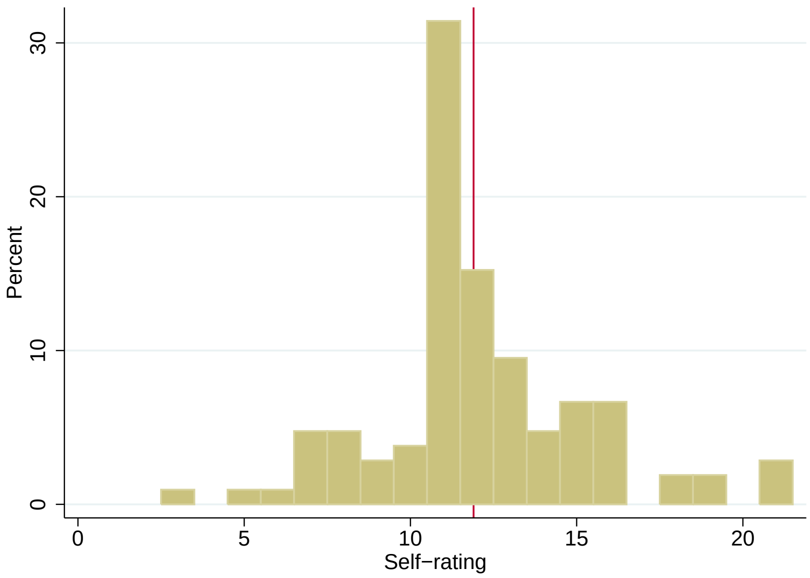

Figure 2: Scatter plot of self-rating and hit rate

This figure displays a scatter plot of self-rating and hit rate. Both measures relate to financial professionals' forecasting skills in foreign exchange. Self-rating is a survey item ranging from 1 to 21 and indicates whether someone believes to be above (21) or below (1) the average hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation.

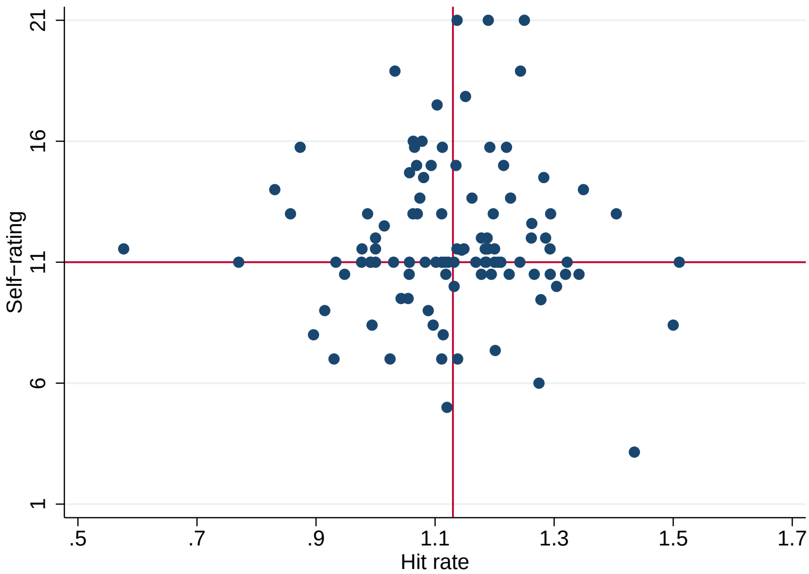

Figure 3: Scatter plot of overconfidence and hit rate

This figure presents the scatter plot of overconfidence and hit rate. Overconfidence is the difference between self-rating and hit rate. Self-rating is a survey item ranging from 1 to 21 and indicates whether someone believes to be above (21) or below (1) the average hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating.

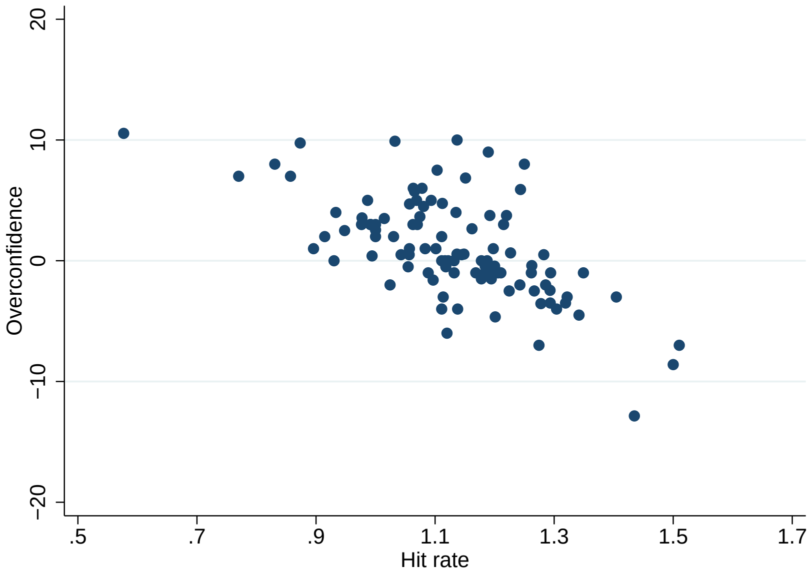

Figure 4: Scatter plot of overconfidence and self-rating

This figure presents the scatter lot of overconfidence and self-rating. Overconfidence is the difference between self-rating and hit rate. Self-rating is a survey item ranging from 1 to 21 and indicates whether someone believes to be above (21) or below (1) the average hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating (see Section 4).

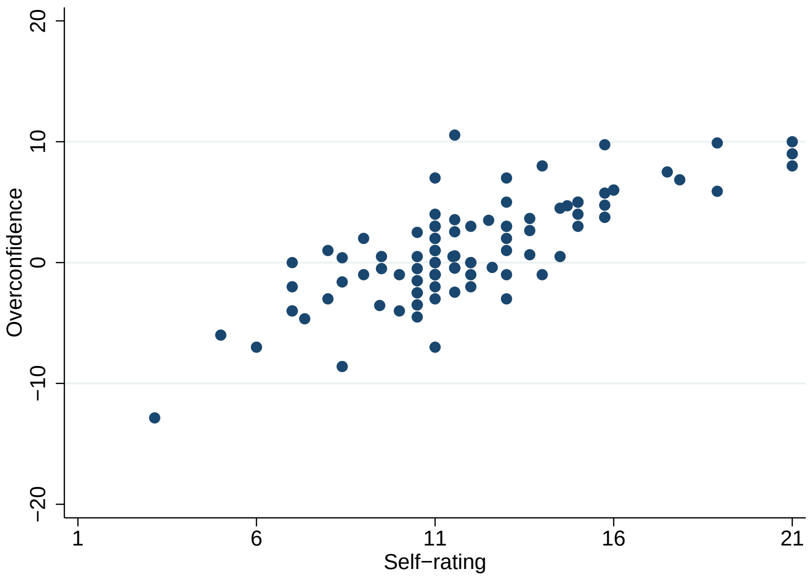

Table 1: Descriptive statistics of overconfidence measures

This table presents descriptive statistics on overconfidence measures of our sample, overconfidence, self-rating, miscalibration, illusion-of.control. To start with self-rating (SR), we ask the financial experts in two surveys (04/2007 and 10/2008) the following question: "How do you evaluate your USD/EUR-forecast compared to the average forecasting hit rate of all participants of the ZEW financial market survey?". The scale ranged from 1 to 21. Overconfidence (OVC) is the difference between self-rating and hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating. The question for miscalibration (MISC) was an item in the survey of October 2008. Respondents gave a 90-% confidence interval for the 6-month future USD/EUR exchange rate. Miscalibration is defined here as the relative confidence interval. Illusion-of-control (IOC) was surveyed in October 2008. The information was extracted from the following question: "Most of the published business news does not surprise me at all." Respondents answered on a scale ranging from 1 to 20. We report Spearman's rank correlation coefficients and the corresponding p-values. The level of significance is denoted by *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$.

								Correlation			
	N	Mean	Median	Std. Dev.	Min	Max		OVC	SR	MISC	IOC
							Corr. coeff	1			
Overconfidence	105	1.13	0.55	4.14	-12.85	10	N	71			
							p-value	.			
							Corr. coeff	0.7704	1		
Self-rating	105	11.91	11	3.15	3.15	21	N	71	71		
							p-value	0***	.		
							Corr. coeff	-0.0446	-0.0444	1	
Miscalibration	74	0.14	0.14	0.07	0.04	0.33	N	71	71	71	
							p-value	0.712	0.7131	.	
							Corr. coeff	0.2623	0.3124	-0.1398	1
Illusion-of-control	73	13.47	14	3.75	4	20	N	71	71	71	71
							p-value	0.0271**	0.008***	0.245	.

Table 2: Descriptive statistics of financial experts' characteristics

This table shows descriptive statistics on demographic, job, and forecasting characteristics of our financial professionals. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Hit rate measures the individual average hit rate over the observation period, where the individual hit rate at a time point codes the forecasting performance in no deviation (2), small deviation (1), and large deviation (0). Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. Dummy variables are denoted by "†".

	N	Mean	Std.Dev.	Min.	Max.
Demographic characteristics					
Male†	105	0.92	0.27	0	1
Academic education†	105	0.76	0.43	0	1
Age	105	44.56	8.11	28.5	64.54
Work experience in fin. sector	105	17.69	8.78	2.5	43.04
Job characteristics					
Operative responsibilities†	105	0.78	0.42	0	1
Personnel responsibilities†	105	0.50	0.50	0	1
Advisor†	105	0.18	0.39	0	1
Fund manager†	105	0.30	0.46	0	1
Researcher†	105	0.23	0.42	0	1
Fundamental analysis	105	55.05	22.41	0	100
Forecasting characteristics					
Hit rate	105	1.13	0.14	0.58	1.51
Positive trend in hit rate†	105	0.15	0.36	0	1
Negative trend in hit rate†	105	0.07	0.25	0	1
Recent success	105	1.18	0.43	0	2
Herding	105	51.42	18.54	1.02	92.08

Table 3: Ordered logit estimation results for overconfidence

This table presents regression results of financial professionals' characteristics on the level of overconfidence. Overconfidence (OVC) is the difference between self-rating and hit rate. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. Dummy variables are denoted by "†". The level of significance is denoted by *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)	(2)	(3)	(4)
Male†	0.436		0.194	
	(0.668)		(0.847)	
Academic education†	-0.249		-0.153	
	(0.682)		(0.787)	
Age	0.0389		0.0464	
	(0.494)		(0.391)	
Work experience in fin. sector	-0.0880*	-0.0578**	-0.0903*	-0.0546**
	(0.100)	(0.0192)	(0.0728)	(0.0285)
Operative responsibility†	0.0546		-0.158	
	(0.940)		(0.829)	
Personnel responsibility†	-0.476		-0.438	
	(0.271)		(0.343)	
Advisor†	0.127		0.0999	
	(0.852)		(0.890)	
Fund Manager†	-1.240**	-1.384***	-1.156**	-1.097***
	(0.0337)	(0.00235)	(0.0329)	(0.00764)
Researcher†	1.023		0.588	
	(0.167)		(0.468)	
Fundamental analysis	-0.0197*	-0.0165*	-0.0152*	
	(0.0517)	(0.0964)	(0.0801)	
Positive trend in hit rate†	1.318*	1.352*		
	(0.0965)	(0.0552)		
Negative trend in hit rate†	-0.763			
	(0.419)			
Recent success			0.814	0.874*
			(0.127)	(0.0519)
Herding	-0.0394***	-0.0305**	-0.0406***	-0.0314**
	(0.00346)	(0.0176)	(0.00369)	(0.0129)
Cut 1	-3.611*	-4.399***	-2.667	-2.554**
	(0.0976)	(0.000168)	(0.185)	(0.0132)
Cut 2	-2.969	-3.785***	-2.035	-1.950**
	(0.168)	(0.000733)	(0.306)	(0.0488)
Pseudo-R2	0.134	0.108	0.121	0.0935
N	105	105	105	105

Table 4: Effects of a marginal/discrete change in the ordered logit regression model

This table displays the change of the level of overconfidence for a marginal/discrete change of the regressors in the fitted ordered logit model. Overconfidence is measured as the difference between self-rating and hit rate. UNC (OVC) corresponds to forecasters who are underconfident (overconfident) and NN to forecasters who are neither nor. The marginal effects are calculated for a change of one standard deviation for the reference case of an average forecaster, i.e. with a working experience of 18 years, about 55 % fundamental analysis usage, herding to the extent of 51 % and who is neither a fund manager nor has a positive trend in the hit rate. The effect for the dummy variables is a discrete change from 0 to 1, denoted by "†".

	UNC	NN	OVC
Work experience in fin. sector	0.1039	0.0197	-0.1236
Fund manager†	0.3296	-0.0085	-0.3212
Fundamental analysis	0.0757	0.0145	-0.0901
Positive trend in hit rate†	-0.1935	-0.0723	0.2658
Herding†	0.1156	0.0219	-0.1374
$P(y\|x)$	28.85	13.96	57.19
$P(y)$	36.19	12.38	51.43

Table 5: Test of different samples

This table displays results of mean comparison tests. Categorical variables are marked by †. We test for differences by χ^2 test statistic and report Fisher's exact p-value. For metric variables we use the t-test. For all variables we report the respective observation number of the sample we use in the analysis and the observations which are neglected in the analysis. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Hit rate measures the individual average hit rate over the observation period, where the individual hit rate at a time point codes the forecasting performance in no deviation (2), small deviation (1), and large deviation (0). The alternative hit rate uses is defined as the binary variable for right (1) and wrong (0) forecasts. Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. The self-rating is a survey item ranging from 1 to 21 and indicates whether someone believes to be above (21) or below (1) the average hit rate. Overconfidence (OVC) is the difference between self-rating and hit rate. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating. The question for miscalibration (MISC) was an item in the survey of October 2008. Respondents gave a 90-% confidence interval for the 6-month future USD/EUR exchange rate. Miscalibration is defined here as the relative confidence interval. Illusion-of-control (IOC) was surveyed in October 2008. The information was extracted from the following question: "Most of the published business news does not surprise me at all." Respondents answered on a scale ranging from 1 to 20. Threshold for fx change is the self-reported appreciation (depreciation) of the exchange rate which corresponds to a change in the qualitative forecast.

Variable	test-statistic	p-value	N (total)	N (in-sample)	N (out-of-sample)
		Panel A			
Male†	0.232	0.622	253	105	148
Academic education†	0.023	1.000	214	105	109
Age	0.996	0.320	202	105	97
Work experience in fin. sector	-0.148	0.882	171	105	66
Operative responsibility†	2.862	0.105	212	105	107
Personnel responsibility†	1.763	0.212	209	105	104
Advisor†	0.718	0.462	220	105	115
Fund manager†	2.669	0.122	220	105	115
Researcher†	1.401	0.307	220	105	115
Fundamental analysis	1.535	0.126	221	105	116
Hit rate	-0.899	0.370	229	105	124
Alternative hit rate	0.905	0.366	229	105	124
Positive trend in hit rate†	1.580	0.230	228	105	123
Negative trend in hit rate†	0.404	0.626	228	105	123
Recent success	0.387	0.699	222	105	117
Herding	1.076	0.283	231	105	126
		Panel B			
Overconfidence†	0.472	0.812	195	105	90
Overconfidence (altern. hit rate)†	2.622	0.278	195	105	90
Self-rating	-0.675	0.500	234	105	129
Miscalibration	-0.154	0.878	164	74	90
Illusion-of-control	-1.621	0.107	162	73	89
Threshold for FX change	1.141	0.257	98	48	50

Table 6: Effect of panel attrition

This table presents the effect of panel attrition on the mean level of overconfidence. We approximate the effect by correlating measures of affiliation to the ZEW panel and overconfidence measures. We report Spearman's rank correlation coefficients and the corresponding p-values. The level of significance is denoted by *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$. We use two measures for the time span of panel affiliation, the number of forecasts during the affiliation and the time span in which a forecaster belongs to the panel (spellduration). The self-rating is a survey item ranging from 1 to 21 and indicates whether someone believes to be above (21) or below (1) the average hit rate. Overconfidence (OVC) is the difference between self-rating and hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating. The question for miscalibration (MISC) was an item in the survey of October 2008. Respondents gave a 90-% confidence interval for the 6-month future USD/EUR exchange rate. Miscalibration is defined here as the relative confidence interval. Illusion-of-control (IOC) was surveyed in October 2008. The information was extracted from the following question: "Most of the published business news does not surprise me at all." Respondents answered on a scale ranging from 1 to 20.

		SR	MISC	IOC	OVC
	Corr. coeff.	0.0438	-0.1461	-0.0487	0.0534
Number of forecasts	p-value	0.657	0.2142	0.6826	0.5883
	N	105	74	73	105
	Corr. coeff.	0.1015	-0.0657	-0.0601	0.1459
Spellduration	p-value	0.3027	0.5781	0.6133	0.1375
	N	105	74	73	105

Table 7: Ordered logit estimation results for overconfidence with alternative hit rate

This table displays regression results of financial professionals' characteristics on the level of overconfidence. Variables referring to hit rate use an alternative coding of right (1) / wrong (0) forecast as a hit rate. Overconfidence (OVC) is the difference between self-rating and hit rate. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. Dummy variables are denoted by "†". The level of significance is denoted by *** p≤0.01, ** p≤0.05, * p≤0.1. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)	(2)	(3)	(4)
Male	-0.186		-1.087	
	(0.835)		(0.331)	
Academic education	0.307		0.491	
	(0.583)		(0.377)	
Age	0.0429		0.0839	
	(0.499)		(0.163)	
Work experience in fin. sector	-0.0971	-0.0674***	-0.109*	-0.0483**
	(0.120)	(0.00520)	(0.0600)	(0.0190)
Operative responsibility	0.577		0.367	
	(0.388)		(0.573)	
Personnel responsibility	-0.716		-0.700	
	(0.137)		(0.150)	
Advisor	0.719		0.570	
	(0.336)		(0.415)	
Fund Manager	-0.352	-0.608	-0.668	-0.813*
	(0.564)	(0.175)	(0.245)	(0.0796)
Researcher	1.673**		0.516	
	(0.0236)		(0.419)	
Fundamental analysis	-0.00345	-0.00169	0.0107	
	(0.755)	(0.866)	(0.300)	
Positive trend in hit rate	2.582**	2.551**		
	(0.0260)	(0.0200)		
Negative trend in hit rate	1.063			
	(0.355)			
Recent success			1.041	1.374*
			(0.212)	(0.0559)
Herding	-0.0509***	-0.0371***	-0.0493***	-0.0383***
	(0.00291)	(0.00841)	(0.000822)	(0.00218)
Cut 1	-2.732	-4.007***	-1.159	-3.207***
	(0.213)	(0.000128)	(0.578)	(0.000131)
Cut 2	-2.008	-3.350***	-0.456	-2.548***
	(0.363)	(0.00138)	(0.828)	(0.00202)
Pseudo-R2	0.196	0.141	0.132	0.0901
N	105	105	105	105

Table 8: Ordered logit results for overconfidence for smaller threshold

This table presents regression results of financial professionals' characteristics on the level of overconfidence. Despite in the base line model we use a threshold of 2% rather than 3% to mark a change of the foreign exchange rate. Overconfidence (OVC) is the difference between self-rating and hit rate. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. Dummy variables are denoted by "†". The level of significance is denoted by *** p≤0.01, ** p≤0.05, * p≤0.1. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)	(2)	(3)	(4)
Male	0.798		0.184	
	(0.459)		(0.852)	
Academic education	-0.470		-0.292	
	(0.433)		(0.597)	
Age	0.0548		0.0520	
	(0.354)		(0.311)	
Work experience in fin. sector	-0.113**	-0.0634**	-0.0949**	-0.0524**
	(0.0432)	(0.0152)	(0.0448)	(0.0296)
Operative responsibility	0.151		0.120	
	(0.842)		(0.872)	
Personnel responsibility	-0.354		-0.272	
	(0.455)		(0.569)	
Advisor	0.438		0.228	
	(0.571)		(0.760)	
Fund Manager	-0.931*	-1.244***	-0.984*	-0.985**
	(0.0830)	(0.00514)	(0.0537)	(0.0219)
Researcher	1.425		0.853	
	(0.100)		(0.321)	
Fundamental analysis	-0.0199**	-0.0155	-0.0152*	
	(0.0363)	(0.100)	(0.0720)	
Positive trend in hit rate	2.015***	1.720***		
	(0.00103)	(0.00128)		
Negative trend in hit rate	-0.136			
	(0.888)			
Recent success			0.576	0.698
			(0.278)	(0.137)
Herding	-0.0527***	-0.0386***	-0.0433***	-0.0331***
	(0.00224)	(0.00254)	(0.00273)	(0.00466)
Cut 1	-3.318	-4.704***	-2.555	-2.721***
	(0.143)	(0.000286)	(0.179)	(0.00870)
Cut 2	-2.959	-4.364***	-2.218	-2.397**
	(0.188)	(0.000564)	(0.241)	(0.0190)
Pseudo-R2	0.160	0.129	0.116	0.0899
N	105	105	105	105

Table 9: Ordered logit results for overconfidence for larger threshold

This table shows regression results of financial professionals' characteristics on the level of overconfidence. Despite in the base line model we use a threshold of 4% rather than 3% to mark a change of the foreign exchange rate. Overconfidence (OVC) is the difference between self-rating and hit rate. Age and work experience in financial sector are measured in years. Fundamental analysis is the self-reported degree (in %) of how much fundamental analysis is used for creating the exchange rate expectation. Positive (negative) trend in hit rate is a dummy variable for a significant positive (negative) trend in the forecasting performance over the last three years. Recent success is the individual average hit rate of 1-month forecasts over the last 6 months. Herding (in %) measures how often a forecaster expects the exchange rate to change in the same direction as the market expected the period before. We refer to market by using the mode of all participating forecasters. Dummy variables are denoted by "†". The level of significance is denoted by *** p≤0.01, ** p≤0.05, * p≤0.1. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)	(2)	(3)	(4)
Male	0.818		0.479	
	(0.409)		(0.637)	
Academic education	-0.248		-0.224	
	(0.672)		(0.679)	
Age	0.00690		0.0243	
	(0.900)		(0.634)	
Work experience in fin. sector	-0.0550	-0.0427*	-0.0635	-0.0501**
	(0.290)	(0.0713)	(0.189)	(0.0429)
Operative responsibility	0.224		-0.0235	
	(0.780)		(0.977)	
Personnel responsibility	-0.485		-0.512	
	(0.269)		(0.266)	
Advisor	0.135		0.0178	
	(0.826)		(0.979)	
Fund Manager	-1.186**	-1.340***	-1.207**	-1.141***
	(0.0423)	(0.00416)	(0.0279)	(0.00579)
Researcher	1.528*		1.232	
	(0.0847)		(0.190)	
Fundamental analysis	-0.0210**	-0.0156	-0.0180**	
	(0.0372)	(0.110)	(0.0474)	
Positive trend in hit rate	1.432	1.470*		
	(0.192)	(0.0973)		
Negative trend in hit rate	-1.022			
	(0.295)			
Recent success			0.543	0.713*
			(0.282)	(0.0843)
Herding	-0.0345**	-0.0241*	-0.0387***	-0.0270**
	(0.0212)	(0.0531)	(0.00644)	(0.0258)
Cut 1	-3.855*	-3.879***	-3.320	-2.551**
	(0.0692)	(0.000586)	(0.103)	(0.0148)
Cut 2	-3.240	-3.309***	-2.721	-1.992**
	(0.122)	(0.00231)	(0.175)	(0.0486)
Pseudo-R2	0.145	0.104	0.127	0.0861
N	105	105	105	105

Table 10: Generalized ordered logit estimation results for overconfidence

This table presents regression results of financial professionals' characteristics on the level of overconfidence. Coefficients are restricted to meet the assumption of proportional odds in that way so that the final model best fits the data. Coefficients give the impact between the respective category and the base outcome (overconfident). Overconfidence (OVC) is the difference between self-rating and hit rate. Dummy variables are denoted by "†". The level of significance is denoted by *** p≤0.01, ** p≤0.05, * p≤0.1. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)		(2)		(3)		(4)	
	UNC	NN	UNC	NN	UNC	NN	UNC	NN
Male	17.32***	-15.75***			18.14***	-16.84***		
	(0)	(0)			(0)	(0)		
Academic education	5.556**	-2.541**			3.428**	-2.649***		
	(0.0226)	(0.0229)			(0.0102)	(0.00480)		
Age	-0.636**	0.136			-0.335*	0.119		
	(0.0388)	(0.428)			(0.0677)	(0.313)		
Work experience in fin. sector	0.598**	-0.202	-0.0474*	-0.0474*	0.280*	-0.257**	-0.0578**	-0.0578**
	(0.0336)	(0.206)	(0.0765)	(0.0765)	(0.0813)	(0.0231)	(0.0240)	(0.0240)
Operative responsibility	2.067*	-1.142			3.236**	-2.063*		
	(0.0586)	(0.148)			(0.0287)	(0.0605)		
Personnel responsibility	-0.493	-0.493			-1.173*	-1.173*		
	(0.405)	(0.405)			(0.0601)	(0.0601)		
Advisor	-2.330*	1.328			-1.768*	1.751**		
	(0.0755)	(0.119)			(0.0592)	(0.0352)		
Fund Manager	-0.977	-0.977	-1.247***	-1.247***	-2.025***	-2.025***	-1.265***	-1.265***
	(0.197)	(0.197)	(0.00785)	(0.00785)	(0.00313)	(0.00313)	(0.00319)	(0.00319)
Researcher	0.511	0.511			2.622*	-1.587		
	(0.454)	(0.454)			(0.0636)	(0.237)		
Fundamental analysis	-0.0334**	-0.000866	-0.0161	-0.0161	-0.0484**	0.00827	-0.0134	-0.0134
	(0.0188)	(0.952)	(0.123)	(0.123)	(0.0282)	(0.585)	(0.127)	(0.127)
Positive trend in hit rate	-21.55***	17.01***	-16.49***	16.33***				
	(0)	(0)	(0)	(0)				
Negative trend in hit rate	1.885	-1.245						
	(0.213)	(0.285)						
Recent success					-0.543	3.330***	0.943**	0.943**
					(0.643)	(0.00194)	(0.0353)	(0.0353)
Herding	-0.0387**	-0.0387**	-0.0333**	-0.0333**	-0.0145	-0.0929***	-0.0331***	-0.0331***
	(0.0377)	(0.0377)	(0.0162)	(0.0162)	(0.503)	(0.00137)	(0.00859)	(0.00859)
Constant	-1.847	18.93***	4.339***	3.653***	-8.856	22.26***	3.400***	2.787**
	(0.814)	(4.56e-06)	(0.000718)	(0.00295)	(0.183)	(1.58e-10)	(0.00324)	(0.0122)
Pseudo-R2	0.431	0.431	0.195	0.195	0.418	0.418	0.104	0.104
N	105	105	105	105	105	105	105	105

Table 11: Ordered logit estimation results for overconfidence (transformation of hit rate by quantiles)
This table presents regression results of financial professionals' characteristics on the level of overconfidence. Overconfidence (OVC) is the difference between self-rating and hit rate. Hit rate is the individual average of the survey forecasts. We code three categories, large deviation (0), small deviation (1) and no deviation (2) of forecast from the true process. Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. For the calculation of overconfidence we rescale the individual hit rate to 1 to 21 to correspond to the range of self-rating. Instead of using 21 equal intervals (as in the benchmark framework) we employ 21 quantiles. Dummy variables are denoted by "†". The level of significance is denoted by *** p≤0.01, ** p≤0.05, * p≤0.1. We report p-values in parenthesis for which we use robust standard errors.

Specification	(1)	(2)	(3)	(4)
Male	1.369		0.488	
	(0.170)		(0.616)	
Academic education	-1.969***		-1.608**	
	(0.00754)		(0.0108)	
Age	0.0393		0.0501	
	(0.546)		(0.324)	
Working experience in fin. sector	-0.111*	-0.0246	-0.0915*	-0.0378
	(0.0885)	(0.289)	(0.0690)	(0.101)
Operative responsibility	0.543		0.237	
	(0.502)		(0.753)	
Personnel responsibility	0.440		0.132	
	(0.430)		(0.795)	
Advisor	-0.799		-0.935	
	(0.300)		(0.210)	
Fund Manager	-1.911***	-1.588***	-1.841***	-1.590***
	(0.00239)	(0.00177)	(0.00197)	(0.00106)
Researcher	1.934*		1.479	
	(0.0881)		(0.169)	
Fundamental analysis	-0.0247**	-0.0113	-0.0159	
	(0.0411)	(0.247)	(0.117)	
Positive trend in hit rate	17.48***	15.71***		
	(0)	(0)		
Negative trend in hit rate	-1.908			
	(0.118)			
Recent success			0.479	0.541
			(0.449)	(0.257)
Herding	-0.0462***	-0.0321**	-0.0483***	-0.0354***
	(0.00947)	(0.0298)	(0.00242)	(0.00983)
Cut 1	-4.264*	-3.432***	-3.661*	-2.809**
	(0.0811)	(0.00359)	(0.0670)	(0.0126)
Cut 2	-4.136*	-3.327***	-3.553*	-2.712**
	(0.0918)	(0.00445)	(0.0758)	(0.0151)
Observations	105	105	105	105
Pseudo-R2	0.311	0.203	0.202	0.129

A Appendix

Figure A.1: Survey question 2007, April

This figure displays the exact wording of the survey questions. US-Dollar/Euro forecasts

1) "How relevant are the following sources of information for your decisions/forecasts (please spend 100 % in total): [ ] % fundamental forecasts (economic and political factors); [ ] % technical analysis (charts, quantitative methods); [ ] % flows (who does what, which orders are in the market)."
2) "How good do you rate yourself compared to a random forecast?" The respondent was supposed to answer on a scale ranging with 21 categories. The lowest category was labeled with "significantly worse", the middle category with "equally", the best category with "significantly better".
3) "How good do you rate yourself compared to the average forecast of the forecaster panel?" The respondent was supposed to answer on a scale ranging with 21 categories. The lowest category was labeled with "significantly worse", the middle category with "equally", the best category with "significantly better".

Sonderfrage: US-Dollar/Euro-Prognosen

1) Welche Relevanz haben folgende Informationsarten für Ihre Entscheidungen/Prognosen (vergeben Sie bitte insgesamt 100%):
   _____ % Fundamentalanalysen (ökonomische und politische Fakten)
   _____ % Technische Analysen (Charts, quantitative Verfahren)
   _____ % Flows (wer macht was, welche Orders liegen vor)

2) Wie gut schätzen Sie die Trefferquote Ihrer US-$/Euro-Prognosen im ZEW-Finanzmarkttest verglichen mit einer *Zufallsprognose*?

Deutlich schlechter — Genauso gut — Deutlich besser

3) Wie gut schätzen Sie Ihre US-$/Euro-Prognosen verglichen mit der *durchschnittlichen Trefferquote* aller Teilnehmer im ZEW-Finanzmarkttest?

Deutlich schlechter — Genauso gut — Deutlich besser

Figure A.2: Survey question 2008, October

This figure displays the exact wording of the survey questions. Exchange rate expectations

1) "Please estimate the USD/EUR exchange rate in 6 months. Please give a range where you expect the exchange rate to be with 90% probability; Lower bound [ ]; Present rate [ ]; Upper bound [ ]."
2) "Most of the published business news does not surprise me at all." Respondents answered on a scale, which was labeled from 1 to 20. The lowest value was labeled with "completely disagree", highest value was labeled with "completely agree".
3) "How good do you rate yourself compared to the average forecast of the forecaster panel?" The respondent was supposed to answer on a scale which was labeled from 1 to 21. The lowest category was labeled with "significantly worse", the middle category with "equally", the best category with "significantly better".

Sonderfrage: Wechselkurserwartungen

1) Bitte schätzen Sie den Stand des US-$/€-Wechselkurs in 6 Monaten ein. Bitte geben Sie dazu eine Spanne an, in der sich der US-$/€-Wechselkurs mit einer Wahrscheinlichkeit von 90% befindet:

Untergrenze: **[** **]** Akt. Stand: **[** **]** Obergrenze: **[** **]**

2) Die Mehrzahl veröffentlichter Wirtschaftsnachrichten stellt für mich keine Überraschung dar.

1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 20
Trifft gar nicht zu Trifft vollkommen zu

3) Wie gut schätzen Sie Ihre US-$/Euro-Prognosen verglichen mit der *durchschnittlichen Trefferquote* aller Teilnehmer im ZEW-Finanzmarkttest?

1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 20 - 21
Deutlich schlechter Genauso gut Deutlich besser

Table A.1: Data sources

This table lists all data sources, which are used in the paper. The exchange rate data originates from the Deutsche Bundesbank, resp. WM/Reuters (accessed via Datastream). All other data is provided by the ZEW from their Financial Market Survey. This survey elicits individual foreign exchange expectations. On top of the regular survey we ask at particular times for additional information. This information includes demographic and job characteristics, three measures from which we derive the level of overconfidence as well as threshold values about what is believed to be an fx change. † The number of observations gives the gross number of respondents. This number does not necessarily equate to the number of observation which we use in our analysis since we clean the original data for inconsistencies.

Data	Source	Date	Observations†
Exchange rate of the USD/DEM	Deutsche Bundesbank	December, 1 1991 to December, 31 1998	na
Exchange rate of the USD/EUR	Datastream	January, 1 1999 to October, 31 2008	na
Expectations of the USD/EUR	ZEW Financial Market Survey	December 1991 to October 2008	na
Demographic characteristics	ZEW Financial Market Survey	September 2003, October 2006	257 (240)
Instruments of fx analysis	ZEW Financial Market Survey	January 2004, April 2007	287 (275)
Better-than-average	ZEW Financial Market Survey	July 2004, October 2008	275 (214)
Miscalibration, illusion-of-control	ZEW Financial Market Survey	October 2008	214
Threshold values of fx change	ZEW Financial Market Survey	August 1997, January 2006	201 (123)

Table A.2: Calculation of hit rates

This table presents the method to compute the hit rates. Hit rates express the performance of exchange rate forecasts. In our study we use monthly forecasts, f_{it}, of the EUR/USD (resp. DEM/USD) exchange rate, S_t. The forecasts are directional forecasts 6-month ahead. The expectation building process can be described as a piece-wise defined function.

$$f_{it}(EUR/USD) = \begin{cases} 1 & \text{if } \frac{E_{it}[S_{t+6m} - S_t]}{S_t} > \varepsilon \\ 2 & \text{if } \varepsilon >= \frac{E_{it}[S_{t+6m} - S_t]}{S_t} >= -\varepsilon \\ 3 & \text{if } -\varepsilon > \frac{E_{it}[S_{t+6m} - S_t]}{S_t} \end{cases} \tag{1}$$

The forecast, f, taking the value one represents an expected appreciation of the USD, 2 equates to a sideways motion, and 3 means an expected USD depreciation relativ to the EUR. ε represents the threshold which corresponds to the deviation of the exchange rate which is believed to be a no change by the forecasters. We know from a survey of the forecaster panel that on average the forecaster associates a change of the exchange rate smaller than 3% with no change in the foreign exchange rate.
Using this information we model a directional times series of foreign exchange changes, d_t.

$$d_t(EUR/USD) = \begin{cases} 1 & \text{if } \frac{S_{t+6m} - S_t}{S_t} > 3\% \\ 2 & \text{if } 3\% >= \frac{S_{t+6m} - S_t}{S_t} >= -3\% \\ 3 & \text{if } -3\% > \frac{S_{t+6m} - S_t}{S_t} \end{cases} \tag{2}$$

Wishing to calculate a precise forecasting performance measure we employ an approach which acknowledges that the experts can choose between three options. A hit rate is coded in three categories: Large deviations are predictions which indicate the opposite direction of the actual movement, whereas small deviations are expectations which are neither a correct forecast nor a large deviation. The code values are 0-1-2 in a way that a higher hit rate is better. This performance measure has been used already earlier for ZEW exchange rate forecasts (e.g. Nolte et al., 2008). Using the series of directional exchange rate changes, d_t, and the individual forecasts, f_{it}, we calculate individual hit rates.

$$HR_{it} = \begin{cases} 0 & \text{if } |F_{it} - D_t| = 2 \\ 1 & \text{if } |F_{it} - D_t| = 1 \\ 2 & \text{if } |F_{it} - D_t| = 0 \end{cases} \tag{3}$$

Calculating the mean over the time for each expert we obtain a precise measure of the true performance of the experts for our analysis. This performance indicator is enhanced by two precautionary measures which we want to emphasize here. First, we consider all forecasts of one person. Second, we exactly determine whether the forecast was right or wrong. In this respect, the survey participants have a time window of about two weeks to submit their forecasts. To achieve a maximum of accuracy and consistency we use individual forecasting days, i.e. we compare the forecasted change of the exchange rate to the realized exchange rate in exactly six months for each individual separately.