Fellner, Gerlinde; Krügel, Sebastian

Working Paper

Overweighting private information: Three measures, one bias?

Jena Economic Research Papers, No. 2010,058

Provided in Cooperation with:
Max Planck Institute of Economics

Suggested Citation: Fellner, Gerlinde; Krügel, Sebastian (2010) : Overweighting private information: Three measures, one bias?, Jena Economic Research Papers, No. 2010,058, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:
https://hdl.handle.net/10419/56809

JENA ECONOMIC RESEARCH PAPERS

2010 – 058

Overweighting Private Information: Three Measures, One Bias?

by

Gerlinde Fellner
Sebastian Krügel

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Jena Economic Research Papers 2010 - 058

Overweighting Private Information: Three Measures, One Bias?*

Gerlinde Fellner† and Sebastian Krügel‡

August 25, 2010

Abstract

Overweighting private information is often used to explain various detrimental decisions. In behavioral economics and finance, it is usually modeled as a direct consequence of misperceiving signal reliability. This bias is typically dubbed overconfidence and linked to the judgment literature in psychology. Empirical tests of the models often fail to find evidence for the predicted effects of overconfidence. These studies assume, however, that a specific type of overconfidence, i.e., "miscalibration," captures the underlying trait. We challenge this assumption and borrow the psychological methodology of single-cue probability learning to obtain a direct measure for overweighting private information. We find that overweighting private information and measures of "miscalibration" are unrelated, indicating that different kinds of misperceptions are at work. Thus, in order to test the theoretical predictions of the overconfidence literature in economics and finance, one cannot rely on the well-established "miscalibration" bias. We find no gender differences in overconfidence for our measures except for one, where women are more overconfident than men.

JEL classification: C91; D03; D83

Keywords: overconfidence; miscalibration; signal perception; cognitive bias

*We are indebted to Martin Beck for research assistance and want to thank the participants of the IMPRS Workshop 2010 in Ringberg and the ESA world meeting 2010 in Copenhagen for valuable comments.

†WU Vienna, Department of Economics, Institute of Economic Policy and Industrial Economics, gfellner@wu.ac.at

‡Max Planck Institute of Economics, International Max Planck Research School 'Uncertainty', kruegel@econ.mpg.de

1 Introduction

Among theorists in behavioral economics and finance, overweighting private information is a bias commonly used to explain some "real world" phenomena. It has been suggested, for instance, as an explanation for the high trading volume observed in financial markets (e.g., Odean, 1998, Kyle and Wang, 1997, Benos, 1998), for the winner's curse in common value auctions (Weyl, 2006), and for short-term market underreactions and long-term market overreactions (Daniel et al., 1998). In these studies, overweighting private information is modeled as a direct consequence of a biased belief about the precision of information (i.e., misperception of signal reliability). That is, an agent who overestimates the precision of his private information, overweights this information when updating his beliefs and therefore acts to his detriment. Overestimating the precision of private information is usually dubbed as "being overconfident," and a link is drawn to a finding in the psychological literature that individuals often overestimate the precision of their knowledge. The latter phenomenon is also called "miscalibration," and it is known to be a specific type of overconfidence. In the economics and finance literature, overconfidence (or miscalibration) is therefore considered to be the underlying trait of individuals who overweight their private information.

Based on the theoretical literature, a few empirical studies have tried to find evidence for the modeled effects of overconfidence (e.g., Biais et al., 2005, Glaser and Weber, 2007). In these studies, overconfidence (or miscalibration) is usually assessed by asking individuals to provide confidence intervals for several knowledge questions. It is well known that in these tasks, individu-

als overestimate the precision of their knowledge such that their confidence intervals are too narrow. Seemingly opposed to the theory, however, miscalibration does not trigger the causal path implied by the above mentioned models: "Measures of miscalibration are, contrary to the predictions of overconfidence models, unrelated to measures of trading volume" (Glaser and Weber, 2007). We argue that such a conclusion is to some extent premature because it heavily relies on the assumption that overestimating the precision of (private) information and overestimating the precision of knowledge together reflect a unitary construct. To the best of our knowledge, this assumption has never been tested, and in view of some related literature it is at least questionable whether it indeed holds.

Several authors in the psychological literature, for instance, have argued in favor of distinguishing between two types of uncertainty: one that is located in the external world and one that is located in the individual himself (e.g., Kahneman and Tversky, 1982, Keren, 1991). When general knowledge questions are asked to obtain measures of miscalibration, the uncertainty is internally located in the individual. Such a task might be more a test of metacognition: "assessors are asked for their knowledge about knowledge" (Keren, 1991). Regarding overweighting of private information, however, the uncertainty is located in the external world, specifically in the information an individual receives. The question is whether an individual misperceives the uncertainty inherent in private information in the same way as he misperceives internal uncertainty when answering knowledge questions.

Beyond this general argument, we also want to point to the literature on forecasting, which clearly distinguishes between three types of forecasting or

judgment tasks based on three types of information (see, e.g., Harvey, 2007): there is (i) forecasting or judgment based on information that is not explicitly available as external data but is held in memory, (ii) forecasting of a variable based on previous values of that variable, and (iii) forecasting of a variable based on explicitly available information about the value of another variable. The bias incorporated in the above mentioned models revolves around the third type of task. However, the bias measured in the empirical studies bears on the first (and second) type of task.[1] Once again, the question remains whether all three types of tasks expose similar personal traits.

This is precisely the question we address in the present study. We rely on a methodology from the literature in psychology on single-cue probability learning (SCPL) to obtain individual measures regarding the weighting of information. In SCPL experiments, subjects predict an outcome based on a single cue over a number of trials. For each subject a prediction slope (subjects' predictions regressed on the corresponding cues) is then compared to the normative slope (true outcome values regressed on the corresponding cues). Subjects who overweight their signals exhibit a prediction slope that is steeper than the normative slope. Thus, using this methodology we are able to obtain a measure that captures the overweighting of information bias directly. Our study is the first to relate this measure to the two measures of miscalibration which are typically employed in the empirical economics literature: miscalibration with respect to general knowledge questions and

[1] Biais et al. (2005) employ only a judgment task of type (i) to obtain a measure of miscalibration. Glaser and Weber (2007) employ tasks of type (i) and (ii) to obtain two measures of miscalibration: one measure based on knowledge questions and the other based on time series predictions. However, neither measure is related to the relevant economic variable (i.e., trading volume in their studies).

miscalibration with respect to time series predictions. If all three tasks rest upon the same underlying trait, individuals who are most miscalibrated in the general knowledge task should also be most miscalibrated in time series predictions, and they should also be the ones who overweight their signals most heavily.

In the next section, we briefly review the relevant literature with an emphasis on the methodology to measure the three, potentially different varieties of judgmental biases. Section 3 and 4 illustrate the design and procedure of our experiment. Section 5 presents the results. In the final section we summarize and discuss our findings and conclude.

2 Relevant Literature

2.1 Overconfidence

Overconfidence has been studied and discussed extensively in the previous literature using different methodologies and various definitions interchangeably. The more recent literature tries to unravel prior inconsistencies and argues that there are several distinct forms of overconfidence. Moore and Healy (2008), for instance, distinguish between overestimation (overestimation of one's ability, performance, and level of control), overplacement (overestimating the relative performance or ability with respect to others (i.e., better-than-average effect), and overprecision (overestimating the accuracy of one's beliefs). Hilton et al. (forthcoming) also suggest that overconfidence may take three forms: judgmental overconfidence (overestimating precision of

one's judgments), self-enhancement biases (positive illusions such as better-than-average effect, illusion of control, unrealistic (personal) optimism), and optimism with respect to societal risks. Miscalibration, the assumed underlying personal trait of individuals regarding overweighting of private information, is a judgment bias and therefore a manifestation of judgmental overconfidence.

General Knowledge Miscalibration

The finding that individuals are often miscalibrated was established by the judgment literature in cognitive psychology (see, e.g., Alpert and Raiffa, 1982, Lichtenstein et al., 1982), which employed two basic approaches to study calibration: individuals are either asked to answer several knowledge questions with two answer alternatives and state their confidence (i.e., their subjective probability) that their answer is correct, or they are asked to construct confidence intervals for knowable magnitudes (e.g., length of a river). Regarding overconfidence in the economics and finance literature, especially the latter method is of interest because it elicits "judgmental overconfidence (...) in a "pure" way" (Hilton et al., forthcoming). When the interval production method is used, the general finding is that individuals' confidence intervals are too narrow: they think they know more about the uncertain quantities than they actually do know. Miscalibration is therefore often defined as an overestimation of the precision of knowledge. It has been found that miscalibration on interval production tasks is a stable personal trait (Hilton et al., forthcoming) with cross-domain consistency (Glaser et al., 2005), and similar results have been obtained for students and professionals

(Glaser et al., 2005). When the interval production method is employed, judgmental overconfidence and overprecision are different labels for the same bias (see the discussion in Hilton et al. (forthcoming) or Moore and Healy (2008)). Following Moore and Healy (2008), we use the term overprecision in the remainder of the paper whenever we refer to miscalibration based on interval production methods.

Time Series Miscalibration

Whereas the tasks in the judgment literature in cognitive psychology are typically comprised of several almanac questions, the tasks in the forecasting literature usually revolve around different types of time series forecasting. The tasks in the forecasting literature therefore differ from the tasks in the calibration literature by the serial correlation of its cues and the presence of history data when a prediction is made (Lawrence and Makridakis, 1989). Some aspects that have been the focus of attention in forecasting research are the influence of data characteristics such as trend, seasonality, and randomness, the influence of the mode of the task presentation (graphical or table format), and the influence of domain-specific knowledge on time series forecasting (a comprehensive review of relevant findings and methodologies can be found in Lawrence et al. (2006)). These aspects have been investigated using different forecasting formats such as point forecasting, probabilistic forecasting, and interval forecasting. Regarding the latter format, it has been found that the above mentioned time series characteristics (i.e., trend, seasonality, and randomness) as well as the presentation scale of the series seem to influence subjects' prediction intervals (e.g., Lawrence and

Makridakis, 1989, Lawrence and O'Connor, 1993, O'Connor and Lawrence, 1992). However, the main finding is – just like in the judgment literature in cognitive psychology – that individuals are overconfident. That is, their prediction intervals are generally too narrow (e.g., Lawrence and Makridakis, 1989, Lawrence and O'Connor, 1993, O'Connor and Lawrence, 1989, Önkal et al., 2003). Interestingly, Glaser et al. (2005) found a positive correlation between overprecision scores based on knowledge questions and overprecision scores based on time series forecasting. This seems to suggest that both tasks indeed share a common underlying trait.

2.2 (Single) Cue Probability Learning

In the psychological analysis of numerical predictions, cue probability learning is the central experimental paradigm. In these experiments, subjects predict an outcome based on a single (or multiple) cue(s) over many rounds. As Ganzach (2009) notes, two approaches have been used to analyze subjects' numerical predictions depending on the research focus. On the one hand, there is the correspondence-based approach of the Social Judgment Theory literature, which focuses on the correlation between the prediction and the true outcome (i.e., achievement index). On the other hand, there is the coherence-based approach, propagated by the Heuristics and Biases program, which compares subjects' predictions against the normative least-square prediction rule: the higher the predictive accuracy of the cue, the higher should be the extremeness of the prediction; the lower the predictive accuracy, the more regressive the predictions ought to be. Because economic

and financial forecasting is concerned with minimizing prediction error, we adopt the coherence-based approach. Hence, the adequate criterion to examine subjects' numerical predictions is the prediction slope.

The main focus of the coherence-based studies lies on aggregated data and on situational factors that influence the extremeness of predictions (e.g., Czaczkes and Ganzach, 1996, Ganzach, 1993, 1994). As a measure of extremeness, these studies typically obtain the ratio of the prediction slopes (i.e., subjects' predictions regressed on the corresponding cues) to the normative slope (i.e., true outcome values regressed on the corresponding cues) and then examine the effect of various feedback and predictor (i.e., cue) representations. In the present study, we adopt the same methodology but use the individual prediction slopes as a measure for subjects' perception of predictive accuracy of their cues. Subjects who overestimate the predictive accuracy will make predictions that are too extreme such that their prediction slopes will be too steep. Consequently, these subjects overweight their signals.[2] Using this measure, we are then able to examine whether there is an empirical relation between overweighting private signals and judgmental overconfidence in interval production tasks based on almanac questions as well as on time series forecasting.

[2]Of course, we do not claim that Bayes' rule is a cognitively valid description of behavior. However, when the individual prediction slopes are used as the yardstick (and therefore Bayes' rule as the benchmark), an individual overweighting his signals is someone who also overestimates the predictive accuracy.

3 Experimental Design

To examine the above mentioned research question, we employ three different judgment tasks based on three types of information and then relate the resulting measures of overconfidence. These tasks form the general design framework of the experiment and are presented in the following.

General Knowledge Questions

The first task is most frequently used in the psychological research on judgmental overconfidence and has thus been taken up by behavioral and experimental research in economics as well (e.g., Biais et al., 2005, Deaves et al., 2008, Glaser and Weber, 2007). It employs the interval production method for general knowledge questions and confronts subjects with almanac questions that require a numerical answer. A question can be of the following kind:

What is the average diameter of the moon (in km)?

As an answer, subjects have to state a lower and an upper bound so that they are 90% sure that the correct answer lies within this interval. Subjects are also instructed that being 90% sure means that for 9 out of 10 questions the true answer should lie within the interval. In our experiment, we asked ten such questions, which can all be found in the Appendix together with the correct answers. A person who is well-calibrated with respect to own knowledge states intervals that contain the correct answer in 9 out of 10 cases. Overconfidence is indicated by intervals that are too narrow, meaning that the correct answer lies outside the subjective confidence interval for

more than 1 out of the 10 questions. Underconfidence, on the other hand, is reflected in too broad confidence intervals so that the correct answer lies within the stated interval in all ten questions.[3] Thus, the number of times the correct answer lies outside the stated interval represents an index for a person's calibration, or more specifically, the estimation of the precision of own knowledge. As a result, the first task produces a general knowledge calibration index ranging from 0 to 10 for each subject, where 1 indicates well-calibration, 0 reflects underconfidence, and increasing numbers reflect higher judgmental overconfidence (i.e., higher overprecision).

Time Series Forecasts

The second task is methodologically similar to the general knowledge task because it also employs the interval production method. However, it is conceptionally different because it aims at subjects' estimation of uncertainty related to time series forecasts. In this task, subjects are presented a time series of an asset value consisting of 20 periods. They are then asked to state an upper and lower bound for the asset value in period 24 so that they are 90% sure the true value will fall within this interval. Again, subjects are told that being 90% sure implies that for 9 out of 10 time series, the true realized value should be within the stated interval.

In our experiment, all time series were based on an autoregressive, moving average process with one MA and one AR term. A trend component was

[3]Since we ask for 90% confidence intervals, there is an obvious asymmetry in the possibility to identify over- and underconfidence. However, we use the same method as many previous studies to be able to relate our results (e.g., Biais et al., 2005, Glaser and Weber, 2007, Hilton et al., forthcoming, Klayman et al., 1999, Russo and Schoemaker, 1992).

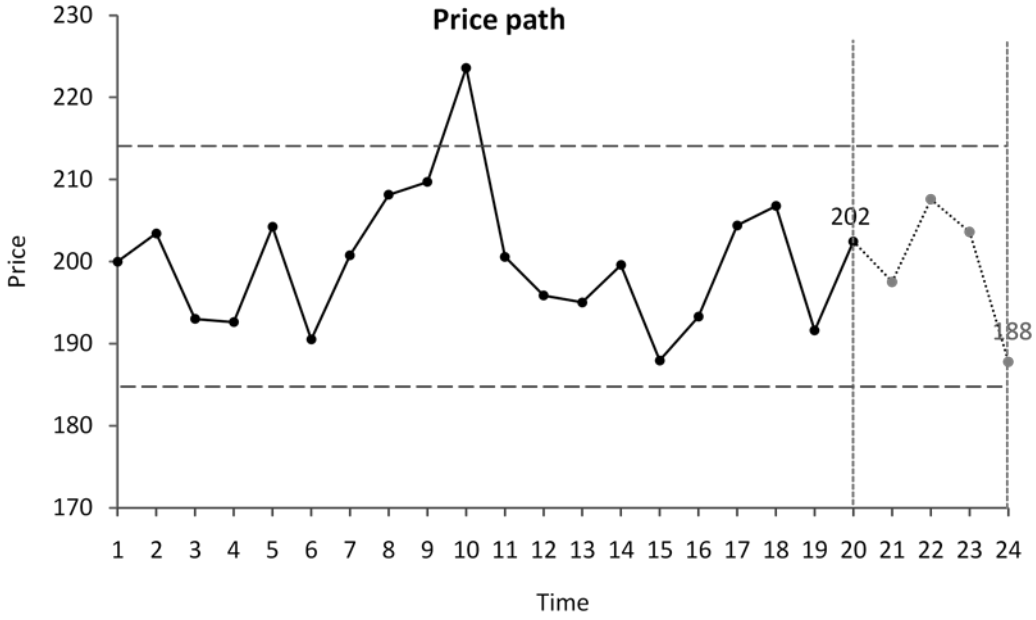

Figure 1: Example for the task of time series forecasts

not included. The so generated time series constitute an "ideal" forecasting environment and have frequently been used in forecasting research (see, e.g., Lawrence and O'Connor, 1992, 1993). We presented ten such time series that were pre-generated using different parameters for the MA and AR term, but all time series had a common starting value of 200. One such example is displayed in Figure 1, where the dashed lines indicate the 90% confidence interval of the realization in period 24. Subjects were, of course, only shown the black solid line of period 1 to 20. All ten time series can be found in the Appendix. The instructions made clear to the subjects that all series were computer generated and that it was therefore impossible to recognize price patterns of real assets. In order to compare whether the true realization of the value in period 24 lay within the stated confidence intervals, the time series were generated for 24 periods of which only the first 20 were presented

to the subjects.

Since the interval method employed here is the same as in the first task, again an individual's calibration index can be obtained but this time with respect to the estimation of uncertainty related to time series forecasting. Thus, similar to the first task, a time series calibration index is calculated as the sum of incidents where the actual asset value in period 24 lies outside the predicted interval. The index ranges from 0 to 10, where 1 indicates well-calibration, 0 reflects underconfidence, and values greater than 1 (increasing) overconfidence (i.e., higher overprecision).

Signal-Based Predictions

While the first two tasks are frequently used to assess subjects' (mis)calibration with respect to knowledge and time series forecasting, our study is the first to relate the obtained measures to a measure of individual signal perception, or, in other words, to the perceived predictive accuracy of a cue. To do so, we strongly rely on the wide field of numerical prediction in the psychological literature, specifically on single-cue probability learning experiments (e.g. Czaczkes and Ganzach, 1996, Ganzach, 1993, 1994). In these experiments, subjects predict an outcome based on a single cue over many rounds, knowing that the cue is a non-perfect, but unbiased indicator of the outcome value. This task perfectly captures subjects' over- or underestimation of the precision of private information: if subjects overestimate the precision of their signal (or cue), they overestimate its predictive accuracy and will therefore make predictions that are closer to the signal than is appropriate. This indicates that the signal is perceived as being too representative of the actual

outcome value and this excessive reliance on private information is referred to as overconfidence in behaviorally inspired economic models (e.g., Odean, 1998, Kyle and Wang, 1997, Benos, 1998).

In order to obtain an individual calibration measure in this task, a prediction slope is calculated for each subject by regressing a subject's predictions on signals. A normative prediction slope can also be calculated by regressing the true outcome values on the signals. If, for instance, the predictive accuracy is overestimated, the information contained in the signal is overweighted, resulting in a prediction slope larger than the normative slope. In general, the steeper the prediction slope (in particular when greater than the normative slope), the higher the overestimation of predictive accuracy.

As in the psychological studies, in our signal-based prediction task subjects have to predict the realization x of a random variable X based on a signal (or cue) s. The random variable X is normally distributed with $N(585, 50^2)$.[4] The signal s, as indicator for x, is determined by $s = x + e$, where e is the realization of a random error term E that is distributed according to $N(0, 50^2)$.[5] The chosen distributional properties of X and E result in a correlation between signal and outcome of 0.7 and in a normative slope of about 0.5.[6]

In a series of 60 rounds, subjects receive a signal s and have to predict x, knowing that the signal is a non-perfect, but unbiased indicator of the value

[4]For reasons of experimental practicality, the distribution was truncated at both ends at four standard deviations so that actual values x were restricted to the range of 385 to 785.

[5]Again, this normal distribution was truncated at both ends at four standard deviations.

[6]We chose these distributional characteristics to be methodologically as close as possible to the studies by Ganzach and coauthors.

x and that there is a positive relation between x and s.[7] For the procedure of the experiment it is important to ensure comparability, so that the values x and e (and thus signal s) for all 60 rounds were pre-generated and kept constant across subjects. Also, x and s were generated with the constraint that the preassigned distributional properties of the two random variables would approximately be preserved within the first, second and third block of 20 rounds. This latter property allows to account for potential learning effects in perceived predictive accuracy.

Two treatments were used for the signal-based prediction task that originate in the two strands of literature we draw upon. The treatments differed only in the degree of prior information about the underlying distribution of the outcome variable. In the *No-Info* treatment, subjects were informed about the range of possible values for x. However, no explicit information was given about the distribution from which value x was drawn. This is the standard procedure used in the studies on single-cue probability learning (e.g., Czaczkes and Ganzach, 1996, Ganzach, 1993, 1994). In the *Info* treatment, on the other hand, subjects were informed about the distributional characteristics of the outcome variable. To ensure an appropriate understanding of the normal distribution, a chart of 1,000 random realizations from the truncated normal distribution was displayed in the instructions (see Appendix). This second treatment was chosen because it more closely captures the overconfidence models in the economic literature: agents are assumed to know the underlying distribution of the central variable (that can be, e.g., the value of an asset). The key information they misjudge is the precision of their private

[7] For more details on instructions, see the Appendix.

signals. So, in the *Info* treatment, subjects received all information except the distributional details on the error variance e of the signal.

Additionally, in both treatments, subjects received a list of ten random draws of x and corresponding signals s prior to starting the task. Thus, they were able to draw some inferences about the predictive accuracy of the signals prior to their first prediction. In each round, they then received the signal and made their prediction of the outcome value at their own pace. After each prediction, the true outcome value x was revealed and they moved on to the next round.

Although each of the three presented tasks aims at detecting patterns of judgmental overconfidence, the underlying cognitive processes might be quite distinct: the general knowledge task uncovers the misperception of own knowledge, the time series prediction task aims at misperception of uncertainty in time series forecasting, and the signal-based prediction task exposes misperception of predictive accuracy of signals. Relating the three calibration measures obtained for each participant will thus clarify whether the synonymous use of these constructs as indicators of judgmental overconfidence can be empirically justified.

4 Procedure

In the experiment, subjects encountered the three tasks in three subsequent phases. Instructions for the general knowledge task (phase one) and the time series forecasts (phase two) were jointly given. In a third phase, the

signal-based prediction task was administered.

It is not easily possible to provide incentive pay for answering general knowledge questions and making time series forecasts. Thus, similar to all other experiments on overconfidence, we paid a flat fee of €3 for finishing the first two tasks. There was no incentive to be very fast in answering the questions because subjects knew they had to wait until everyone was ready to start with the next phase.

To provide incentives for the repetitive signal-based prediction task, one out of the 60 rounds was randomly selected for payment. In this task subjects earned a flat fee of €6, which was reduced according to the absolute deviation of their prediction from the true value x. For every integer of deviation, €0.015 were subtracted from the fee of €6. To facilitate understanding of this compensation scheme, the instructions contained a payoff table with a number of examples.[8]

Participants in the experiment were 168 students from Jena University, 85 females and 83 males. In the third phase, 88 subjects were assigned to the Info-treatment, 80 to the No-Info treatment. Subjects were recruited using the software tool ORSEE (Greiner, 2004), and the experiment was conducted with the software z-Tree (Fischbacher, 2007). The three phases of the experiment lasted about 1 hour, and average earnings accumulated over

[8]In principle, a quadratic scoring rule has the preferable property of incentive compatibility under the assumption that subjects are risk neutral. Such an assumption is, however, challenged by the findings of risk attitude elicitation, which was conducted in a later phase of the experiment. Moreover, the quadratic scoring rule is difficult to understand for participants and thus likely overburdens an otherwise rather simple decision. Sonnemans and Offerman (2001) show that subjects do not exhibit less effort in making good decisions when being paid with a flat fee instead of a quadratic scoring rule.

these phases amounted to €10.93 including a show-up fee of €2.5.[9]

5 Results

First, we report how participants are classified according to the accuracy of their judgments in the three tasks and present an overview of average overconfidence. Subsequently, the relation of the three overconfidence measures is examined while controlling for other influencing factors. Finally, we look for gender effects that are frequently reported in studies on judgmental overconfidence.

5.1 Descriptive Overview of Overconfidence Measures

Table 1 provides a descriptive overview of the measures of judgmental overconfidence in all three tasks. Columns (2) and (3) contain the data obtained from the general knowledge questions and the time series forecasting. In both of these tasks, subjects were asked to state 90% confidence intervals to ten questions each. For both tasks, a calibration index is calculated as the number of times the true values fall outside the stated intervals. As mentioned above, a well-calibrated individual should have an index of 1.

Overconfidence in General Knowledge

The mean calibration index for the general knowledge task is 5.8, indicating

[9] After completing these stages, subjects participated in two further stages, consisting of a risk attitude elicitation task and an experimental asset market, which lasted for another 1.5 hours and earned them an additional €12.21, on average. The results of these stages are reported in a different paper.

considerable overconfidence. Thus, on average, for 58% of the questions the correct answers fall outside subjects' confidence intervals. This is well in line with prior studies.[10] Moreover, Table 1 indicates large individual differences in this task with calibration indices ranging from 0 to 10. The overwhelming majority of our subjects (159 or 94.6%) is overconfident with an index ranging from 2 to 10. Only 5 subjects (or 3.0%) are well-calibrated and 4 subjects (or 2.4%) are underconfident. The Cronbach's alpha for the general knowledge calibration index is 0.69, indicating an acceptable psychometric validity.

Overconfidence in Time Series Forecasting

The mean calibration index for the time series forecasting task is 1.2, therefore indicating only slight overall overconfidence in our sample and being much less than the average calibration index obtained in the general knowledge task. Moreover, the classification of our subjects based on the time series calibration index is much more balanced: about equally many subjects are overconfident (45 or 26.8%) and well-calibrated (43 or 25.6%); 80 subjects (or 47.6%) are underconfident. However, it is not the aim of this study to examine whether and why different degrees of overconfidence exist between the general knowledge and the time series forecasting task. Rather, we are interested in the question whether individuals who are most overconfident in one task are also most overconfident in another. Thus, we only need a ranking of our subjects with respect to all overconfidence measures, not the

[10] Russo and Schoemaker (1992), e.g., find a percentage of correct answers falling outside the stated confidence intervals in the range from 42% to 64%. In Hilton et al. (forthcoming), the percentage of answers outside the intervals is between 62% and 78%, and in a study by Glaser and Weber (2007), the percentage of answers outside the intervals is 75%.

actual degree of overconfidence. It therefore suffices that there are individual differences within each task. And indeed, even though the individual differences are not as pronounced as in the general knowledge task, they also exist with respect to time series forecasting. The calibration index for this task ranges from 0 to 7.[11] The Cronbach's alpha for the time series calibration index is 0.74, once again indicating an acceptable psychometric validity.

Table 1: Miscalibration measures

	General knowledge	Time series forecasts	Signal-based predictions	
			Info	No Info
Well Calibrated	1	1	0.5	0.5
No. obs.	168	168	88	80
Mean score (SD)	5.8 (2.4)	1.2 (1.6)	0.65 (0.16)	0.87 (0.14)
Min	0	0	0.36	0.37
Max	10	7	1.02	1.06
# overconfident	159 (94.6%)	45 (26.8%)	65 (73.9%)	77 (96.2%)
# well calibrated	5 (3.0%)	43 (25.6%)	13 (14.8%)	2 (2.5%)
# underconfident	4 (2.4%)	80 (47.6%)	10 (11.4%)	1 (1.3%)

Overconfidence in Signal-Based Predictions

Columns (4) and (5) of Table 1 contain the data obtained from the signal-based prediction task. In this task, subjects were asked to predict an outcome

[11] Yet one might argue that a ranking based on the time series calibration index is too crude since, e.g., 80 of our subjects are underconfident with a calibration index of 0. However, the use of generated time series has an additional advantage in this respect. Based on a similar methodology as in Lawrence and O'Connor (1993), we randomly sampled 100 possible outcomes for the value of each time series in period 24. This enables us to calculate a much finer calibration index for each subject ranging from 0 to 1000. When this procedure is used, a calibration score of 100 indicates perfect calibration. Nonetheless, using this calibration index for the time series forecasting task does not change any of the later results qualitatively. Thus, for reasons of simplicity we present all the results based on the above described calibration index for the time series forecasting task.

based on a signal over 60 rounds, knowing that the signal is a non-perfect but unbiased indicator of the outcome value. In the *Info* treatment (column 4) subjects were informed about the characteristics of the underlying distribution of the outcome value. In the *No-Info* treatment (column 5), they did not receive this information. As a measure of perceived predictive accuracy of the signals, we calculated a prediction slope for each subject by regressing the predictions in all 60 rounds on the corresponding signals. The normative slope which minimizes the prediction error can be obtained by regressing the true outcome values on the signals. The normative slope in this task is approximately 0.5. Thus, to adhere to the prior labeling, we call an individual with a prediction slope of 0.5 well-calibrated; a prediction slope greater than 0.5 indicates overestimation of predictive accuracy and thus overconfidence; a prediction slope smaller than 0.5 indicates underestimation of predictive accuracy and thus underconfidence.[12]

In both treatments, the mean prediction slope is greater than 0.5, indicating an overestimation of the predictive accuracy of the signals overall. With an average prediction slope of 0.65, this bias is significantly less pronounced in the *Info* treatment compared to an average prediction slope of 0.87 in the *No-Info* treatment (Wilcoxon rank sum test: $p < .01$). Thus, even though subjects in the *Info* treatment overweight their signals on average, they are closer to the normative prediction slope because they know the underlying distribution of the outcome value. Subjects in the *No-Info* treatment, on the other hand, were not informed about the distribution from which the out-

[12]In order to classify an individual as well-calibrated with respect to signal perception, we set a range for the prediction slope by tolerating a deviation from the normative slope of ± 0.05.

come value was drawn. In addition to overestimating the precision of their signals, they might therefore also overestimate the variability (i.e., variance) of the underlying distribution of the outcome value. Both processes lead to an overestimation of the predictive accuracy of their signals and thus, to a greater overweighting bias.

In both treatments large individual differences are prevalent. In the *Info* treatment the prediction slope ranges from 0.36 to 1.02, and in the *No-Info* treatment it ranges from 0.37 to 1.06. Accordingly, in the *Info* treatment 65 subjects (or 73.9%) are classified as overconfident, 13 subjects (or 14.8%) as well-calibrated, and 10 subjects (or 11.4%) as underconfident. In the *No-Info* treatment, on the other hand, the overwhelming majority (77 subjects or 96.2%) is overconfident, 2 subjects (or 2.5%) are well-calibrated, and only 1 subject (or 1.3%) is underconfident.

The dynamic nature of the signal-based prediction task allows to investigate whether perception of predictive accuracy changes with task experience. Figure 2 gives an overview of prediction slopes when the total number of prediction rounds is split into three blocks of 20 rounds. In the *Info* treatment, the average prediction slope decreases significantly over time (all pairwise comparisons are significant with $p < .01$ according to Wilcoxon signed ranks tests), indicating that subjects become better calibrated in the perception of signal precision in later rounds. In the *No-Info* treatment, on the other hand, calibration becomes only slightly better. The average prediction slope does not differ between the first and second block and between the second and third block ($p = .32$ and $p = .30$, respectively). Comparing blocks 1 and 3 reveals significantly better calibration in the last block ($p = .03$).

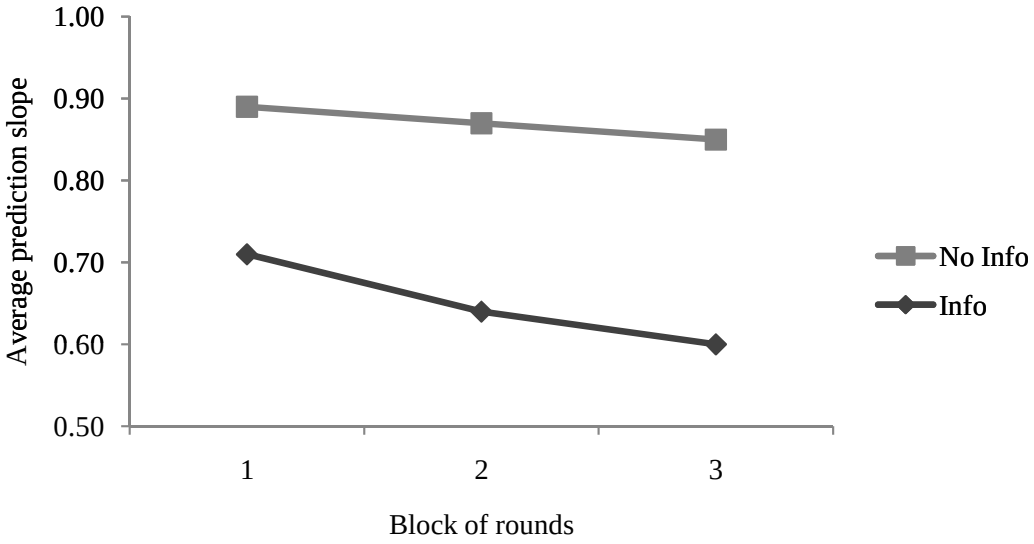

Figure 2: Accuracy in signal-based predictions over time

With an average prediction slope of 0.71 in the *Info* treatment compared to 0.89 in the *No-Info* treatment, the difference between the two treatments is already significant in the first block of 20 rounds (Wilcoxon rank sum test: $p < 0.01$). In summary, this suggests that knowledge about the underlying distribution of the outcome value does not only facilitate learning how to be better calibrated with respect to the perception of predictive accuracy but also improves the initial calibration.

Before we can investigate the individual stability of overconfidence across tasks, we have to examine whether the rank ordering of our subjects in the signal-based prediction task is stable. Because learning is involved in the task, subjects who highly overestimate the predictive accuracy of the signals in the first rounds, for example, might improve their predictions more than other subjects in the later rounds and vice versa. If this is the case, the rank ordering of our subjects would differ between different stages of the

task. To address this issue, we again split the total number of prediction rounds into three blocks of 20 and calculate the prediction slope for each subject in each block. Based on these prediction slopes, we then compute all pairwise Spearman correlation coefficients between the three blocks as well as the Spearman correlations between each block and the prediction slopes based on all 60 rounds. Table 2 contains the results. Even though there is an overall learning effect in the signal-based prediction task, stable individual differences are prevalent across the blocks. All pairwise correlations between blocks are positive and significant, ranging from 0.65 to 0.84 for the *Info* treatment and from 0.45 to 0.73 for the *No-Info* treatment. Thus, subjects who are relative overweighters of signals at the beginning of the task are likely to be relative overweighters at the end of the task as well. Moreover, the prediction slopes of each block are highly correlated with the prediction slopes based on all 60 rounds, ranging from 0.88 to 0.94 in the *Info* treatment and from 0.73 to 0.88 in the *No-Info* treatment. This suggests that the average slope based on the predictions in all 60 rounds is a good summary statistic regarding overweighting of private information for each subject.

5.2 Relation of Overconfidence Measures

5.2.1 General Knowledge and Time Series Forecasting

Table 3 shows the correlation matrix of overconfidence scores based on the general knowledge and time series forecasting task. In addition to the common miscalibration measure, the correlation matrix also includes an interval

Table 2: Spearman correlations between prediction slopes

Treatment	Slopes	Block 1	Block 2	Block 3	Total
Info	Block 1	–			
	Block 2	0.74***	–		
	Block 3	0.65***	0.84***	–	
	Total	0.88***	0.94***	0.90***	–
No Info	Block 1	–			
	Block 2	0.52***	–		
	Block 3	0.45***	0.73***	–	
	Total	0.73***	0.87***	0.88***	–

*** significant at 0.01

width and accuracy score for both tasks. These measures allow for a more detailed assessment of miscalibration. Subjects' interval width scores are calculated by ranking the interval width across participants for each item and summing the ranks for each subject across the ten questions of each task. Thus, the higher a subject's interval width score, the wider his confidence intervals tend to be, relative to those of the other subjects. The accuracy score is obtained by the same procedure, but instead of the interval width, the ranking of subjects is now based on the absolute distance between the midpoint of the stated interval and the true answer. Thus, the higher a subject's accuracy score, the farther away his midpoints tend to be from the true answers.

As expected, a tendency to use wider intervals in the general knowledge and time series forecasting task is related to lower miscalibration scores in each task ($r = -0.70$, $p < 0.01$ and $r = -0.79$, $p < 0.01$, respectively). Similarly, the farther the midpoints from the true answers, the higher the miscalibration score in each task ($r = 0.25$, $p < 0.01$ and $r = 0.26$, $p < 0.01$, respectively). Interestingly, we also find a positive and significant correlation

between the interval width and the accuracy score in the general knowledge task ($r = 0.22$, $p < 0.01$), which indicates that the farther away the interval midpoints from the true answer, the wider the intervals tend to be. This suggests that subjects rightly react to a higher degree of uncertainty by widening their confidence intervals. However, a widening of the intervals is not sufficient to adjust for the greater deviation of their midpoints, as the positive correlation between the accuracy score and the miscalibration measure indicates.

Most importantly for our study, we find a positive correlation between the two miscalibration measures ($r = .50$, $p < 0.01$). Subjects with a high miscalibration score in the general knowledge task tend to be the subjects with a high miscalibration score in the time series task. This suggests that both measures capture a common construct which is in line with prior studies (see, e.g., Glaser et al., 2005). Additionally, the interval width score based on the general knowledge task is significantly correlated with the miscalibration score based on the time series forecasting task ($r = -.54$, $p < 0.01$) and vice versa ($r = -.57$, $p < 0.01$). Thus, subjects who state narrower intervals in one task also have a higher miscalibration score in the other. This, too, suggests that both miscalibration tasks measure a common construct, which is based on a general tendency to use narrow intervals. We therefore generalize the results of Hilton et al. (forthcoming), who find a significant correlation between interval width and miscalibration scores across two different general knowledge scales.

Table 3: Spearman correlations between different miscalibration measures

			Scores based on Knowledge		Scores based on Time Series	
	Knowledge	Time Series	I.Width	Acc.Sc.	I.Width	Acc.Sc.
Miscalibration Measures						
Knowledge	–					
Time Series	0.50***	–				
Scores based on Knowledge						
Interval Width	−0.70***	−0.54***	–			
Accuracy Score	0.25***	0.04	0.22***	–		
Scores based on Time Series						
Interval Width	−0.57***	−0.79***	0.63***	0.04	–	
Accuracy Score	0.00	0.26***	−0.03	−0.01	−0.01	–

Note: Correlation coefficients are based on all data (n=168). *** indicates significance at 0.01.

5.2.2 Correlations with Signal-Based Predictions

Table 4 shows the Spearman correlations between different measures of overconfidence based on the first two tasks and the prediction slopes, separately for treatment *Info* and *No-Info*. Columns (1) and (2) contain simple pairwise correlations while columns (3) to (6) contain partial correlations. For the partial correlations we control for the influence of gender, age, and semester as well as the remaining measures of miscalibration for which a correlation coefficient is reported in each particular column. Thus, the correlations between the general knowledge miscalibration score and prediction slope in columns (3) and (5), for instance, are the partial correlations between both measures while controlling for gender, age, semester, and time series miscalibration. Regarding the correlations between both of these measures in columns (4) and (6), we additionally control for the interval width scores in the general knowledge and time series forecasting task.

As shown in the table, prediction slopes and measures of miscalibration in the general knowledge task are not correlated. Between time series mis-

calibration and the prediction slopes, we find a positive and significant correlation in the *Info* treatment only. This suggests that in the *Info* treatment participants with a higher miscalibration score in the time series forecasting task tend to overweight their signals more heavily in the signal-based prediction task. The correlation coefficient is rather small, though (between .26 and .31). Moreover, when controlling for interval width, the correlation between time series miscalibration and the prediction slope in the *Info* treatment becomes only marginally significant ($r = .19$, $p = .085$). Partial correlations between interval width and prediction slopes are never significant. Thus, subjects who tend to state narrower intervals in the general knowledge or time series forecasting task do not systematically overweight their signals more heavily. This is surprising as the interval width score corresponds most closely to Moore and Healy's (2008) "overprecision," the type of overconfidence usually assumed to be the underlying trait of individuals who overweight their private information. In general, the correlation results in Table 4 suggest that judgmental overconfidence assessed through confidence interval production methods, on the one hand, and assessed through overweighting of externally given signals, on the other, are two distinct constructs.

5.3 Gender Differences in Overconfidence

Finally, we investigate whether the degree of overconfidence differs between men and women, as claimed by some previous literature (e.g., Barber and Odean, 2001). Table 5 contains means and medians of our overconfidence

Table 4: Spearman correlations between measures of miscalibration and prediction slopes

	Simple Correlations		Partial Correlations			
	Info	No-Info	Info		No-Info	
	(1) Slope	(2) Slope	(3) Slope	(4) Slope	(5) Slope	(6) Slope
Measures based on Knowledge						
Miscalibration	0.17	0.09	-0.03	-0.06	0.14	0.18
	(0.109)	(0.409)	(0.786)	(0.602)	(0.237)	(0.112)
Interval Width	–	–	–	-0.05	–	0.09
				(0.625)		(0.460)
Measures based on Time Series						
Miscalibration	0.31***	-0.05	0.26**	0.19*	-0.10	0.02
	(0.003)	(0.644)	(0.015)	(0.085)	(0.384)	(0.869)
Interval Width	–	–	–	0.01	–	0.09
				(0.895)		(0.446)

Note: For the partial Spearman correlations other additional control variables not included in the table are gender, age, and semester. *** significant at 0.01, ** significant at 0.05, * significant at 0.10

measures separately for men and women as well as the number of observations in each task. The last column of Table 5 contains the p-values of a Mann-Whitney U test, where the null hypothesis is equality of populations. We found a significant difference of average overconfidence between men and women in the general knowledge task only. In this task, women have a higher miscalibration score than men (6.18 vs. 5.43, $p = 0.03$). However, as the comparison of the interval width score indicates, this is not due to a general tendency of women to use narrower intervals than men. Rather, the higher miscalibration of women in this task can be ascribed to a lower accuracy of their judgments, as the higher value of the accuracy score indicates (965.11 vs. 804.37, $p < 0.01$). Thus, on average, women's interval midpoints are farther away from the correct answers than the interval midpoints of men. In this sense, given their knowledge, women in our sample are more overconfident than men in the general knowledge task. Regarding overconfidence measures

Table 5: Gender differences in several measures of miscalibration

	Male			Female			p-value
	Mean	Median	Obs	Mean	Median	Obs	
Measures based on Knowledge			83				
Miscalibration	5.43	5	83	6.18	7	85	0.03**
Interval Width	875.85	881.5	83	894.91	879.5	85	0.93
Accuracy Score	804.37	803	83	965.11	961	85	0.00***
Measures based on Time Series							
Miscalibration	1.05	0	83	1.26	1	85	0.52
Interval Width	913.89	869.5	83	876.26	892	85	0.62
Accuracy Score	889.36	868.5	83	867.39	851.5	85	0.68
Prediction slopes							
Treatment Info	0.62	0.62	43	0.67	0.64	45	0.15
Treatment No-Info	0.87	0.94	40	0.87	0.88	40	0.32

*** significant at 0.01, ** significant at 0.05. p-values are based on Mann-Whitney U tests.

based on the time series forecasting task as well as the signal-based prediction task, we do not find significant gender differences.

6 Summary and Discussion

In behavioral economics and finance, overweighting private information is a bias that is often used as an explanation for empirical phenomena of detrimental decision making, like the winner's curse or strategies of excessive trading. In theoretical modeling, it is usually captured by (overconfident) agents who overestimate the precision of their private signals. In empirical tests of these models, however, it is generally assumed that the modeled bias resembles a specific type of overconfidence identified in the calibration literature in cognitive psychology where individuals are asked to state confidence intervals for general knowledge questions. We have put this assumption to the test. Based on the psychological literature on forecasting, we argue that different cognitive mechanisms might be triggered in tasks involving uncer-

tainty that is located internally (like assessing the precision of own knowledge in almanac questions) and in tasks involving uncertainty that is located externally (like assessing the precision of signals about an asset value). The lack of empirical support for economic models of overconfidence (e.g., in the area of trading) might thus originate in divergent empirical and theoretical constructs of judgmental overconfidence.

We employ three types of judgment tasks and investigate whether overconfidence measures obtained in these tasks are correlated. The first two tasks are established in the overconfidence literature and require to state subjective confidence intervals for answers to general knowledge questions and time series forecasts. We introduce a third type of task, signal-based predictions, which borrows the methodology from the psychological literature on single-cue probability learning. In so doing, we obtain a measure for overweighting private signals that closely resembles overconfidence in the way it is captured in economic models. If the assumption of one underlying personal trait in all three tasks holds, individuals who are most overconfident in one task should also be most overconfident in the other tasks.

Similar to the previous literature, we find, on average, substantial overconfidence in the general knowledge task and, to a lesser degree, also in the time series forecasting task. In the signal-based prediction task, overconfidence is also prevalent overall. However, we observe a lower degree of overconfidence when subjects know about the distribution of the outcome variable they have to predict (*Info* treatment) than when they do not know about this distribution (*No-Info* treatment). In the latter treatment, two possible sources for overestimating predictive accuracy of signals come into

consideration: overestimating signal precision (as in the *Info* treatment) or overestimating the variability of the outcome distribution (or both). Yet, notwithstanding the exact sources of overestimating the predictive accuracy of signals, it is important to note that both treatments uncover overweighting of signals (i.e., overconfidence in signal perception).

Regarding gender differences, we cannot confirm the often raised claim that men are more overconfident than women (e.g., Barber and Odean, 2001). For most of our overconfidence measures we do not find any significant gender differences, which is in line with some prior studies (see, e.g., Gigerenzer et al., 1991). In the general knowledge task, however, we find that women are more miscalibrated than men. This is not due to a general tendency of women to use narrower intervals than men but to a lower accuracy of their judgments. Women therefore seem to be more overconfident than men in the general knowledge task.

Relating overconfidence across the judgmental tasks reveals that measures of miscalibration based on a general knowledge and time series forecasting task are positively and significantly correlated. Thus, subjects who tend to be most overconfident in the general knowledge task tend to be most overconfident in the time series forecasting task as well. Moreover, we find that interval width scores are correlated with miscalibration measures across the two different tasks, suggesting that miscalibration is due to a general tendency to use narrow intervals. In sum, those results indicate that both tasks indeed uncover one underlying personal trait, which closely corresponds to Moore and Healy's (2008) "overprecision" (i.e., narrow intervals).

With respect to signal-based predictions, however, the assumption of one

underlying trait cannot be maintained. Overweighting of signals and measures of miscalibration seem to be unrelated. Only in one of our two treatments for signal-based predictions do we find a significant correlation between overweighting of signals and time series miscalibration. The correlation is rather small, though, and after controlling for interval width, it becomes only marginally significant. Moreover, interval width scores are not significantly correlated with measures regarding the weighting of signals. This is particularly surprising as the interval width scores most closely capture the type of overconfidence assumed to be the underlying trait of individuals who overweight their private information. In general, this suggests that overweighting of private information and overconfidence assessed through confidence interval production methods are two distinct biases. Hence, we conclude that a discrepancy exists between modeling and measuring overconfidence in the economic literature.

References

Alpert, M., Raiffa, H., 1982. A progress report on the training of probability assessors. Judgment under uncertainty: Heuristics and biases, 294–305.

Barber, B., Odean, T., 2001. Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment. Quarterly Journal of Economics 116 (1), 261–292.

Benos, A., 1998. Aggressiveness and survival of overconfident traders. Journal of Financial Markets 1, 353–383.

Biais, B., Hilton, D., Mazurier, K., 2005. Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. The Review of economic studies 72 (2), 287–312.

Czaczkes, B., Ganzach, Y., 1996. The natural selection of prediction heuristics: Anchoring and adjustment versus representativeness. Journal of Behavioral Decision Making 9 (2), 125–139.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839–1886.

Deaves, R., Luders, E., Luo, G., 2008. An experimental test of the impact of overconfidence and gender on trading activity. Review of Finance.

Fischbacher, U., 2007. Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

Ganzach, Y., 1993. Predictor representation and prediction strategies. Organizational Behavior and Human Decision Processes 56, 190–190.

Ganzach, Y., 1994. Feedback representation and prediction strategies. Organizational Behavior and Human Decision Processes 59, 391–391.

Ganzach, Y., 2009. Coherence and correspondence in the psychological analysis of numerical predictions: How error-prone heuristics are replaced by ecologically valid heuristics. Judgment and Decision Making 4 (2), 175–185.

Gigerenzer, G., Hoffrage, U., Kleinbölting, H., 1991. Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review 98 (4), 506–528.

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

Glaser, M., Weber, M., Langer, T., 2005. Overconfidence of professionals and lay men: Individual differences within and between tasks? University of Mannheim, Working Paper Series SFB 504 (05-25).

Greiner, B., 2004. An online recruitment system for economic experiments. In: Kremer, K., Macho, V. (Eds.), Forschung und wissenschaftliches Rechnen, GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, pp. 79–83.

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

Hilton, D., Régner, I., Cabantous, L., Charalambides, L., Vautier, S., forthcoming. Do positive illusions predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration. Journal of Behavioral Decision Making.

Kahneman, D., Tversky, A., 1982. Variants of uncertainty. Cognition 11 (2), 143–157.

Keren, G., 1991. Calibration and probability judgements: conceptual and methodological issues. Acta Psychologica 77 (3), 217–273.

Klayman, J., Soll, J., González-Vallejo, C., Barlas, S., 1999. Overconfidence: It depends on how, what, and whom you ask. Organizational Behavior and Human Decision Processes 79 (3), 216–247.

Kyle, A., Wang, F., 1997. Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? Journal of Finance 52, 2073–2090.

Lawrence, M., Goodwin, P., O'Connor, M., Önkal, D., 2006. Judgmental forecasting: A review of progress over the last 25years. International Journal of Forecasting 22 (3), 493–518.

Lawrence, M., Makridakis, S., 1989. Factors affecting judgmental forecasts and confidence intervals. Organizational Behavior and Human Decision Processes 43 (2), 172–187.

Lawrence, M., O'Connor, M., 1992. Exploring judgemental forecasting. International Journal of Forecasting 8 (1), 15–26.

Lawrence, M., O'Connor, M., 1993. Scale, variability, and the calibration of judgmental prediction intervals. Organizational behavior and human decision processes(Print) 56 (3), 441–458.

Lichtenstein, S., Fischhoff, B., Phillips, L., 1982. Calibration and probabilities: The state of the art to l980. Judgment under uncertainty: Heuristics and biases, 306–334.

Moore, D., Healy, P., 2008. The trouble with overconfidence. Psychological review 115 (2), 502–517.

Önkal, D., Yates, J., Simga-Mugan, C., Oeztin, S., 2003. Professional vs. amateur judgment accuracy: The case of foreign exchange rates. Organizational Behavior and Human Decision Processes 91 (2), 169–185.

O'Connor, M., Lawrence, M., 1989. An examination of the accuracy of judgemental confidence intervals in time series forecasting. Journal of Forecasting 8 (2), 141–155.

O'Connor, M., Lawrence, M., 1992. Time series characteristics and the widths of judgemental confidence intervals. International Journal of Forecasting 7 (4), 413–420.

Odean, T., 1998. Volume, volatility, price and profit: when all traders are above average. Journal of Finance 53, 1887–1934.

Russo, J., Schoemaker, P., 1992. Managing overconfidence. Sloan Management Review 33 (2), 7–17.

Sonnemans, J., Offerman, T., 2001. Is the quadratic scoring rule really incentive compatible? Working PaperUniversity of Amsterdam.

Weyl, E., 2006. Biasing auctions. Unpublished Manuscript.

A General knowledge questions

The following ten questions were used in the general knowledge task. Correct answers are in parentheses.

1. What is the length of the river Nile in km? (6,671 km)
2. How many states are currently *(Nov. 2009)* members of the OPEC? (12)
3. What is the average diameter of the moon in km? (3,745 km)
4. What was the number of inhabitants of Australia in 2008 (in Mill.)? (21.374 Mill.)
5. What is the number of passenger airports in Germany? (38)
6. What was the number of patent applications in Germany in 2008? (62,417)
7. What is the size of France in km^2? (674,843 km^2)
8. What is the air distance between London and Tokio in km? (9,581 km)
9. When was the novel Robinson Crusoe by Daniel Defoe first published? (1719)
10. When was the zip fastener patent-registered? (1893)

B Time series forecasting task

The following ten time series to be used in the time series forecasting task were pre-generated using an autoregressive, moving average process with one MA and one AR term.