Hindsight bias, risk perception and investment performance

Bruno Blikas (Edinburgh University) & Martin Weber (Munich University)

June 2006

Abstract

Once they have observed information, hindsight biased agents fail to remember how ignorant they were initially - they knew it all along! We formulate a theoretical model of this bias, providing a foundation for empirical measures and implying that hindsight biased agents learning about volatility will underestimate it. In an experiment involving 67 students from Munich University, we find that hindsight bias reduces volatility estimates. In another experiment involving 85 investment bankers in London and Frankfurt, we find that more biased approaches have lower performance. These effects are robust to differences in location, education, confidence and experience.

Many thanks to the participants in the Edinburgh and Amsterdam University seminars, the Warwick Conference on Risk and Finance (2005), the Munich Summer School on Economics and Psychology (2005), and the UU conference on biased beliefs, in particular Denis Halton, Jaki Gusten, Robin Hogarth, Florentino Lopez-Simon, Thomas Mirrlees, Frank Veldman, Jan Tastle and Richard Zeckhauser. We like to thank Alex Nourie for excellent research assistance.

1
1 Introduction

Psychologists have extensively documented the prevalence of cognitive biases and deviations from rational decision. Are these results relevant for actual financial markets? Professionals operating in their natural habitat and within a lot of noise could be expected to behave as理性ly. Biases could be worked out by learning and experience. Market forces and economic environments could induce optimism. The present paper combines experimental evidence from the lab and from the field to shed light on these issues.

Decision making in financial markets relies crucially on information processing and learning. Efficient learning requires comparing new information to previous expectations. For example, after earnings announcements, investors must compare the news to their prior expectations. They must take into account the information context (e.g., 'Is this a shock or the norm?'). The hindsight bias, which is the inability to correctly remember one's prior expectations after observing new information, hinders such information processing. Hindsight bias is not surprising, as new information is what the scene on the balance. The present paper is therefore concerned with the question of whether the bias does not eliminate.

The aim of this paper is to study the consequences of the hindsight bias for an important type of economic activity: investment and trading. We propose a simple theoretical model providing

a formulation for the measurement of the bias and showing that hindsight biased agents fail to assess variances correctly. When volatility is stochastic, traders need to update their assessment of risks based on return realizations. On observing unexpectedly positive or negative returns, rational agents should make their volatility estimates. Hindsight biased agents who “know it all along,” fail to understand that such returns were unexpected and thus underestimate variances. More generally, hindsight biased agents will form inaccurate beliefs about asset returns, leading to suboptimal trades and inferior financial performance. For example, hindsight biased agents underestimate volatility will underestimate options and fail to hedge appropriately. Hindsight biased agents will also fail to call in question economic analyses at odds with the facts. And they will fail to estimate accurately the difference between their information and that of other traders. This will undermine their ability to earn trading profits based on superior information.

To test the hypothesis that the hindsight bias hindered learning about risks, we designed a lab experiment involving 67 students from Münster University. We gave the participants specific data and asked them to estimate volatilities. Then, we gave them new data and asked them to again estimate volatilities. The idea was to study how participants process the new data to update their volatility estimates. The experiment involved two treatments. In the first treatment, the experimenters reminded the participants of their initial estimates, thus making them biased. In the second one, the experimenters asked the participants to remember their initial estimates so that the bias could manifest itself. Consistently with our theoretical predictions, in second first treatment participants came up with lower variance estimates than in the first one.

To test the hypothesis that the hindsight bias reduces performance, we collected data from 85 investment bankers working in Frankfurt or London. We found these bankers to be significantly hindsight biased. And yet the question we asked them related to their own field of expertise, such as stock market returns, mean variances or characteristics of the investment banking industry. We also found that more experienced bankers or bankers with more precise
information were not less biased. Using data provided by the bank, we predicted bankers’ perfor-
mination by their compensation (including bonuses). Consistent with the hypothesis that the
highlights biases versus performance, we found that the bankers in the highest earning category
had the lowest bias on average. We then expanded this finding in the Frankfurt sample. To test
its robustness, we collected additional data for London. In this new sample we found the
same result. We also checked that our results were not driven by other variables such as the
consumption expenditure or information of the bankers.

Our paper is in line with Camerer et al. (1993). We use similar methods to measure the
highlights bias, but our focus and findings differ from theirs. Camerer et al. (1993) showed that
markets reduce but don’t eliminate the highlights bias. We complement this finding by offering
new results on the consequences of the bias for information processing and financial performance.

Another distinguishing feature of our paper is that our findings on the link between the highlights
bias and performance are based on data collected in the field, from highly paid and experienced
investment bankers. This is in the line of several recent field experiments on financial decisions
and psychology. Our analysis is also in the line of several recent papers empirically relating
psychological mechanisms to economic variables (see Camerer (1995), Prelec (1998),
Christie et al. (2000), and Ghader and Weber (2007)).

In the next section we present our theoretical model. In Section 3, we present our lab
experiments. In Section 4, we present our field experiment. Section 5 concludes.

We sample H Noah and Lat (2010) empirically find that managers are more sensitive for CEO
excess returns, while Sutton & Schleifer (1994) finds the highlights on investment bankers
prices is also prevalent among investment bankers.
2 Theory

2.1 Defining the hindsight bias

The research on hindsight bias started with Fishbein (1975) and Fishbein and Beatty (1975). Afterwards have been noticed, that people are more likely to remember their erroneous expectations. Hindsight bias suggests that, by a process of retroactive reconstruction, they incorporate the new knowledge into their recollection of the prior expectation (see e.g., Carroll, 1999). In the work of Wänke et al. (1994, p. 403) this is a projection of emotional attachment to the past expectation by a denial that the current information has influenced judgments. Correspondingly, for hindsight biased agents, the retrospective reconstruction of the initial belief will be closer to the realization than the true ex-ante expectation. Numerous papers in experimental psychology have shown the presence of this bias and discussed its measurement (see e.g., Wänke and Häsler, 2000).

March (2006) found empirically that individual differences in hindsight bias exist. This suggests that the bias acts as a trait, consistently influencing individual behavior in various environments. The notion of hindsight bias was initially developed in the context of binary variables. In that case, the bias arises if the expected recollection of the ex-ante probability is greater when the event actually occurred. The bias arising in the general case is often referred to as the “error of knowledge.” In the present paper, we will not differentiate cases of knowledge and hindsight bias.

Let \( y \) be a random variable with ex-ante expectation \( \mu \). Suppose the agent is told the realization \( y \) and asked to report his ex-post expectation. Internally, he reports \( \mu \). If hindsight bias is for him to remember his ignorance, he was initially and forms a biased reconstruction of the prior mean, tilted towards the realization of the random variable. This can be modeled by specifying the reconstructed prior mean of the biased agent as a weighted average of the true
\[ E(\omega) = \omega \mu + (1 - \omega) \bar{\mu} \]

where the constant \( \omega \) measures the magnitude of the bias and \( \bar{\mu} \) denotes the biased expectation.

The subscript denotes that the biased expectation is reported after observing the realization \( v \) of \( \mathcal{N} \). This formulation is similar to equation (1) in Chasco et al. (2006). When applied to experts, the prior mean of the distributions natural experts are not influenced by the realization of the random variable so that \( \omega = 1 \). In contrast, biased experts partially forget the ex-ante expectations. If they are completely biased, they totally forget the natural prior mean and place all the weight on the realization of the random variable (\( \omega = 1 \)).

2.2 Adverse consequences of the hindsight bias

Several previous papers have discussed adverse consequences of the hindsight bias. For example, Mungas and Weber (1998) and Madaras (2008) show that, in a principal agent relation, the hindsight bias will prevent the principal from correctly evaluating the performance of the agent. In other words, principals fail to remember what was known when the agent's decision was taken, which is in line with the finding of Bar on and Horvitz (1998). Iller et al. (2013) find that, when asked questions about outcomes before and after the introduction of the RIAS, RIAS supports showed stronger hindsight bias for positive developments, while opponents showed stronger bias for negative developments.

One of the main disadvantages of the hindsight bias is that it prevents rational processing of information and learning from past data. Thus, Hershfield (1982) notes that the hindsight bias will prevent one from rejecting one's hypotheses about the world because of hindsight.
we systematically underestimate the priors that the past holds for us, we are subjecting these hypotheses to uninnately weak tests and presumably finding

Reason to change theme.“ (Field and Glick, 1998, page 344)

Similarly, observing students analyzing business cases, Biheller and Goolly (1988) found that

the hindsight bias has hindered learning from past experience.

In financial markets the inability to learn from the past and to reject hypotheses can be very damaging. Hindsight biased traders will fail to recognize that their view of the market was wrong. Hence they will fail to cut their losses when it is optimal to do so. Hindsight biased investors will incorrectly take into account the informational content of new signals such as announcements or micro-sees. This will lead them to form suboptimal portfolios. In the remainder of this section we focus on the adverse consequences of the hindsight bias for risk assessment.

Risk is one of the most important elements of the environment of financial decisions. Correctly factoring risk in these decisions is made difficult by random fluctuations in volatility. Because of these fluctuations, it is important for agents to conduct efficient learning about volatility. The hindsight bias hampers such efficient learning. To see this consider the following simple model. With probability $\lambda$ the random variable $V$ is normally distributed with variance $\sigma_1^2$, and with probability $1-\lambda$ it is normal with variance $\sigma_2^2$, where $\sigma_2 > \sigma_1$. After observing one realization of $V$, the agent has to update her beliefs about the true variance. Applying Bayes law, rational agents would update the probability that the variance is lower to

$$\lambda \left( \frac{\sigma_2}{\sigma_1} \right)^2 \left( \frac{\lambda}{1-\lambda} \right) = \lambda \left( \frac{\sigma_2}{\sigma_1} \right)^2 \left( \frac{\lambda}{1-\lambda} \right).$$

Then the

[The previous version of this paper included the final analysis of such institutions possible for the reason, see Kline and Weber (2007).]
(2) \[ P(\varphi) = \frac{1}{1 + e^{-x/\theta}} \]

Hindsight bias agents will proceed in a similar updating except that, instead of using \( E[P] = \mu \), they will rely on the biased expectations \( P(\varphi) = \mu + (1 - \mu) \). This leads to the following biased probability:

(3) \[ P(\varphi) = \frac{1}{1 + e^{(\varphi - \mu)/\theta}} \]

Combining (2) and (3), we obtain the following proposition:

**Proposition 2:** In our framework, when agents must conduct learning about volatility, those with hindsight bias underestimate volatility.

Agents increase the probability that volatility is high when they observe realizations that differ a lot from their prior expectations, i.e., when they are surprised. Hindsight biased agents tend to not be surprised: “It could all happen” hence, they underestimate volatility.

Underestimation of risk induced by the hindsight bias will adversely affect decision making. Hindsight biased traders will underestimate the value of options and hedging strategies. They will incorrectly assess risk and return underestimation. They will incorrectly estimate risk premium. Hence, the hindsight bias reduces performance.

### 2.3 Measuring the hindsight bias

There are different empirical designs that have been used to measure the hindsight bias.

Fishhoff and Beyth (1975) use a within person design where subjects are first asked to report their current expectations. Then, they learn the realization of the variable and are asked to report their current recollection of their current expectations.
Fischhoff (1975) complements this design by adding a control group. At the stage where subjects are asked to report their predictions of their initial estimate two groups are formed. In one group, they receive no information. In the other group, they are told the true outcome of the event.

Camerer et al. (1993) use a third design also with two groups. In group one subjects are told the true outcome of the event and asked to estimate the average expectation of group one (knowing that this group had no information).

To pursue the way for our empirical analysis in the next section, we now analyze our simple model of the mechanism corresponding to design 1 and 2. Before being told the answer to a question, people don’t know it exactly. For them it is a random variable. In line with the notation of our simple model we denote it by $X$. For simplicity we assume information is homogenous.

In this context, first consider the “within person design” of Fischhoff and Betts (1977). In that design, the hindsight bias can be measured by the ratio of $i)$ the agent’s remembered estimate ($E_i(x)$) using his initial estimate ($\mu$) to $ii)$ the difference between the realization of the variable ($x$) and the agent’s initial estimate ($\mu$). Substituting $E_i(x)$ from equation (1), this ratio simplifies to $1 - \frac{x}{\mu}$.

Next, turn to the “between subjects design” of Camerer et al. (1993). They form two groups (A and B) and ask a question to the participants in group A. Denote by $\bar{X}_A$ the average answer of group A participants. The experimenter gives the answer $x$ to participants in group B. Then, the experimenter asks these participants to assess the average answer of group B. Denote by $\hat{P}(W_B)$ the assessment of $W_B$ by participant $c$ in group $A$. The index of hindsight bias used by Camerer et al. (1993) is

$$\frac{\hat{P}(W_B)|W_B=x}{1 - \frac{x}{\mu}}$$

The previous section of this paper included an extension of this analysis to the case where agents observe private signals (Babcock and Odoni, 1993).
The difference between the individual's estimate of the average group B's answer minus the actual average, divided by the difference between the true answer and group B's average answer. As noted by Chewning et al. (2008), unbiased agents should on average correctly estimate $W_B$, hence the average hindsight bias index is 0. In contrast, biased agents will be influenced by the revelation of $n$. If completely biased, they will believe the others gave the exact right answers and the index will be 1. If partially biased, they will believe

$$\hat{E}(W_B) = \mu_B + (1 - \beta)\mu$$

Noting that all participants in group B should give the same answer $\mu$ so that $W_B = \mu$, and substituting $W_B$ and $\hat{E}(W_B)$ in equation (4), we obtain our second Proposition.

**Proposition 2**: In our framework, the Chewning et al. (2008) index for each participant is equal to his/her hindsight bias parameter ($\beta$).

### 3 Study 1: hindsight bias and risk perception

The goal of this experiment is to test whether hindsight bias affects the way people perceive how much uncertainty there was ex ante, and thus come up with lower risk estimates than unbiased agents. The basic setup has two treatments. In the first treatment, the hindsight bias is created ex ante. In the second treatment, the hindsight bias is created ex post by the experimenter informing the participants of their ex ante estimates. In the first treatment, the hindsight bias manipulation has an effect because the participants are not told their initial forecasts, neither are asked to remember these forecasts. We test the hypothesis that participants give higher volatility estimates in the first treatment than in the second one.

#### 3.1 Experimental design

The experiment was run in the context of the behavioral finance class taught by one of the authors (Martin Weber) to fourth-year students in Mannheim University. 67 students participated.
in the experiment, which involved two tasks.

The first task was on 29 November 2005. We gave the participants the market price of 5
Commodities (Wheat, Rice, Soy, Palm oil, and Sugar) and 2 commodities (oil and gold) as
well as the euro-dollar exchange rate for 21 and 28 November. We asked them to predict the price one week later (on December 5). We also asked them to give an upper bound and a lower bound such that there would be only one chance out of ten that the December 5 price would be outside the bounds. We then transformed this confidence interval into a standard
deviation using the method introduced by Boxer and Bodily (1989). Thus, we directly elicited
their estimate of the expected price and indirectly estimated their estimate of the standard
deviation. We noticed that such an indirect elicitation because it is more natural and intuitive
for participants to give a confidence interval than a standard deviation.

The second task was a week later, on December 6. At that point we gave the participants
the realized price for December 5 (and also reminded them of the November 28 price). We asked
them again to give us our predictions (for December 5) expectations and confidence intervals
for the 5 stocks, the 2 commodities, and the exchange rate.

Students were randomly assigned to one of two treatments. In the first treatment, on Decem-
ber 6, we gave the participants their estimates (prediction and bounds) from last week (formed
on November 28 for December 5). In the second treatment we did not remind the participants
of their estimates from the previous week. Rather we asked them to remember and select
these estimates (prediction and bounds). Thus, while in the first treatment, reminded the block-
sight bias by explicitly reminding the participants of their forecasts in the second treatment
the hindsight bias could manifest itself. 35 students participated in the first treatment, 32 par-
ticipated in the second treatment. At each round we randomly drew 10 participants who each
received 10 euros for participation.1

1 Probably this amount financial incentive would not be big enough to influence studies assumed.
Our empirical results are summarized in Table 1. First consider the hindsight bias. As mentioned above, the number of...
munities are presented in the second row of Table 1 (along with, in parentheses, the values for
denominating that the ratio is equal to 1. For 5 out of 8 the median ratio is significantly
above 1 at the 10\% level. The median ratio is strictly below 1 for only one event (and in that
case the difference to 1 is not statistically significant). This also suggests that participants are
blindlight biased on average. As far as we can tell, this is a new way to document this bias.

As discussed in the previous section, blindlight biased agents are “never surprised”. In our
experiments, in line with the model presented in the previous section, we measure surprise as the
square of the difference between the prediction and the expectation, divided by the standard
deviation. When this surprise is measured using the initial expectation and volatility estimates
of the agents, we refer to it as the “true surprise.” When it is measured using the remembered
expectation and volatility estimates, we refer to it as the “biased surprise.” Of course, while
the true surprise can be computed in both treatments, the biased surprise can be computed
in the second treatment only. The biased surprise is a way to jointly assess the effect of the
blindlight bias on expectations and on stock prices. For each agent and each event, we computed
the true surprise and in the second treatment, the biased surprise. We then took the median
across agents. These medians are depicted in Table 1, in the third, fourth and fifth rows. As
expected, there is no systematic ranking between the true surprises between the two treatments,
and the differences are not significantly different from 0 (the p-values are above 10\%) for all 8
events. In contrast, for the second treatment, the biased surprise is systematically smaller than
the true surprise. However, the difference is significant in only two cases (44-4 and 44-5).

The main objective of this study was to test the implication from our theoretical analysis
that the inability to be surprised leads biased agents to come up with lower volatility estimates
than unbiased agents (see Proposition 1). To document this point we computed, for each agent
and each participant, the ratio of the initial one-period ahead volatility estimate (formed on

\footnote{The figures have been rounded to integers.}
November 29) to the subsequent computation about volatility estimate (formed on December 6). We then took the medians across agents. These medians are depicted in the last two rows of Table 1. For all assets except CDS, the median ratio is lower for the first treatment, and in 4 cases the difference is statistically significant. Thus, the spread revision of volatility forecasts is stronger for participants who were shown their initial estimates than for participants who are asked to remember initial estimates. This is consistent with the hypothesis that hindsight bias against remembered estimates leads to lower volatility updates, because they are never surprised.

Finally, note that our design offers the opportunity to observe two variables related to the hindsight bias. On one hand, we measure (a) the ratio of the difference between the agent’s remembered estimate and his initial estimate to (b) the difference between the revisions of the variable and the agent’s initial estimate. On the other hand, people with stronger hindsight bias are predicted to have a greater revision of initial implied volatility to remembered implied volatility. For each of the 8 assets in our study, we computed the correlation across participants between these two variables. The Spearman correlation coefficients (and corresponding p-values) are: 0.38 (0.05) for BAS, -0.06 (0.7) for BAC, 0.43 (0.01) for CBO, 0.28 (0.19) for Del'haize, 0.43 (0.08) for Educor, 0.27 (0.12) for Goldman, 0.47 (0.06) for Google, and 0.25 (0.22) for Unilever. Consistent with our theory, the vast majority of the coefficients is positive. But the result is not very strong, since the correlation is significantly different from 0 only in one case (0.4).

4 Study 2: hindsight bias and performance

While the goal of Study 1 was to analyze the effects of the hindsight bias on the way in which agents process information, the goal of Study 2 is to test its effects on performance.
4.1 Data Collection

To assess the effect of the self-kitten bias on performance in the field we collected psychometric and financial data for bankers working in a large investment bank, in Helsinki and London. The psychometric data used in the present study was collected as part of a larger seven-page questionnaire where other psychological measures were also assessed. In the present study, we use data related to the self-kitten bias and also to control for other possible effects, the misattribution bias and the better than average effect.

The participants in the study are from the trading department of the bank. While they have various tasks (sales, research and trading), they all participate in the elaboration of portfolio allocations, trading and investment decisions. The data was collected by inviting groups of participants into one of two conference rooms in the bank. There, they filled out the questionnaire under the supervision of an experimenter. After filling the questionnaires which took about 30 minutes, participants were asked not to talk to their colleagues until the end of the data collection. The overall data collection effort took about two hours. The data was collected anonymously with each questionnaire being tied with a number. This number was used to sort the questionnaires into earnings categories. For each number the personnel department of the bank informed us whether the respondent was in the top, middle or low earnings category. Earnings meant overall compensation, including bonuses. Setting subjects into these categories was a compromise between our wishes (as much performance data as possible) and the privacy requirements of the participants and institution. It should be noted that the distribution of earnings assessed (as suggested by common wisdom) and an interview with the human resource department). Therefore earnings in the low earnings and medium classes are less variable than in the high earnings class.

*These variables, together with another one collected from another task, are analyzed in Chiesi et al. (2013).*
We first collected the data in the Frankfurt branch of the bank, on December 3, 2001. We had 41 respondents (each member of the department who was present that day). Thus to ensure the integrity of our results, we collected similar data in the London branch of the bank on October 5, 2002. There we had 49 respondents, corresponding to the vast majority of the department. Bankers in both locations had similar jobs. In Frankfurt, the questionnaires had to be approved by the management and by the worker's council. We made it very explicit that the study was done by two universities and that the management of the bank would not see any specific questionnaires. The only data made available to participants and management were aggregate data. No payments were given.

4.2 Descriptive statistics

Table 2 offers some descriptive statistics on the population of bankers in our sample. It shows that low and high-earnings bankers have approximately the same age in Frankfurt and London. High-earnings bankers are considerably older in London. In Frankfurt, the bankers are relatively equally distributed across the different departments (sales, trading, research and others). Unfortunately, the London sample did not fill in our questionnaire. Hence, the bankers in our London sample only belong to the sales department.

To estimate the hindsight biases of the investment bankers, we used the “between-person” experimental design of Cameron et al. (1999). We found that highly competitive bankers would find it difficult to admit they could not perfectly forecast the future. Such consideration could have contaminated hindsight bias measures obtained in a “within-person” design (see, e.g., Campbell and Tesser 1991). Also, as Balfe (1999) points out, as economists, we can pretend a

*The estimates based on questionnaires, however, only distinguish two groups of earning categories. Hence, to make the treatment comparable, the use of a questionnaire is tested on mixed groups of employees and bank employees. The results are based on 120 respondents.
Our experimental design involved two types of tasks for each participant. In the first task, they were asked to estimate the true value of some unknown variables (true estimate). In the second task, they had to guess the average answers of the other group to some questions (false estimate). In that second task, the participants were given the true answer to the question asked to the other group. Each subject had to answer ten questions, five in the true estimate condition and five in the false estimate condition. We formed two groups in each location. Former referred to as Group A and B.

The items were taken from the natural habitat of the bears. Nevertheless, we used questions for which we thought they did not know exactly the correct answer. The items were similar for Transilvania and London. Whenever it was reasonable, the questions were exactly the same for the two locations. The second sample however, was collected 10 months later than the first and in a different country. Hence, we had to change some questions to ensure comparability. The questions are listed in Table 1.

Table 1 reports the average hindsight bias, measured using the Camerer et al. (1997) index. Participants in our study did not show a clear trend in hindsight bias as they did not know the natural habitat of the bears and that the participants are experienced professionals. The degree of hindsight bias is different for different questions but between 0 and 1 in the majority of cases. And the magnitude of the hindsight bias was observed in this between-person design is similar to that observed in the within-person design of Study 1. To summarize the data we computed the average bias across participants and across the 5 questions. The mean hindsight bias is very similar in London (247) and in Transilvania (250). It is significantly lower than 1 and about 0 in both locations. The t-statistics are 5.00 and 4.25 respectively for London and 1.41 and 2.2 respectively for Transilvania.
4.3 Psychometric issues

4.3.1 Precision of information

Does our measure of the hindsight bias really capture this bias, or spuriously reflect other variables? One possibility is that it simply mirrors the precision of the knowledge of the participants regarding the questions we asked them. To check this we estimated a measure of this precision. In the overestimate treatment, for each participant and each question, we computed the absolute value of the difference between the participant’s answer and the true answer. Then we summed these mistakes across questions. Finally, we normalized the total mistake of each participant by the average in his group. The greater this ratio, which we hereafter refer to as the mistake ratio, the lower the precision of the information of the participant. To examine if the hindsight bias is related to the mistake ratio, we computed the correlation between the two. The Spearman correlation coefficient is 0.066. The probability to observe such a value under the null that the two variables are uncorrelated is 41%. Hence the mistake ratio and the hindsight bias are weakly and insignificantly negatively correlated.

4.3.2 Underconfidence and overconfidence

Another possibility is that our measure of the hindsight bias actually reflects underconfidence. The agent would believe the others to be better informed than him and thus they really are. In this context, while the agent would be surprised by the realization of the random variable, he/she would believe that the others had predicted it very accurately and were not surprised. While this theory of underconfidence could generate positive values for the Correct-it (1995) index, in practice this is unlikely to be driving the results in our experiment. As mentioned above, in our questionnaire we collected the answer to “better than average” questions. Participants were asked to answer what percentage of their colleagues they expected to perform better than
themselves along four dimensions (trading skills, communication skills, market vision, technical skills). Averaging across the four dimensions, and across participants, the bankers in our sample answered that 34.6% of the others had better skills than themselves. This suggests overconfidence rather than underconfidence. In addition, the correlation across bankers of the better than average score with the hindsight bias index is not significantly different from 0. For the Frankfurt bankers it is 0.005 (with a p-value of 0.790). For the London bankers the correlation is 0.007 (p-value = 0.744).

4.3.3 Miscalibration

Another form of overconfidence which has received attention in finance is miscalibration (see Biais et al. (2005)). In line with Hsu and Schoemaker (1992) and Kipman et al. (1999), we measure miscalibration by clicking confidence intervals. We show participants ten questions. For each question we ask them to give an upper bound and a lower bound such that the true answer should lie outside the bounds with only one chance in ten. If participants were correctly calibrated the frequency of answers lying outside the bounds should be 10%. In our sample, the frequency of such answers was 69%. Such strong miscalibration is consistent with the findings of Kipman et al. (1999) and Hansen and Schoemaker (1992). In line with this literature we then use miscalibration to assess the percentage of questions for which the true answer lies outside the indicated bounds. To assess if miscalibration is independent from the hindsight bias, we computed the correlation between the miscalibration score and the hindsight bias score. For the London sample this correlation is equal to 0.3, while for the Frankfurt sample it is equal to 0.4. In both samples, the correlation is not significantly different from zero.
4.4 Empirical results

Up to this point we have demonstrated that investment bankers exhibit hindsight bias. We have
found that our estimate of hindsight is not significantly correlated with other variables such as
information precision and overconfidence. But does the hindsight bias affect performance? We
now turn to that question.

4.4.1 Main result

As can be seen from Table 4, the hindsight bias (median as well as mean) is lower for the high
earnings category than for the two other categories. The average hindsight bias is somewhat
lower for the low earnings category than for the middle earnings category. Note however that
when one focuses on median the difference is not large. Table 4 also shows that, for the middle
earnings category, the median bias is much lower than the mean, while for the high earnings
category, the median bias is much lower than the mean, and for the low earnings category, there
is no clear pattern. This suggests that the mean in the high earnings category is driven down by
a few participants with very low bias, while the mean in the middle category is driven up by a
few participants with very high bias.

In addition, Table 4 shows that the relation between the hindsight bias of bankers and their
performance is present in both locations. Both in London and in Frankfurt the bankers with
the highest performance were those with the lowest bias. This robustness speaks to the cross
of-sample validity of the results. It is particularly striking since, as was mentioned above, we first
designed and implemented the experiment with Frankfurt participants, and then to evaluate
the robustness of our results we replicated the same experiment with London participants.

Table 4 also suggests that the result is not driven by other variables such as age or experience.
Indeed, the age and experience structure differ in the two locations. While in London the high
earnings bankers have longer experience than the middle earnings bankers (18 versus 14), in
To evaluate the significance of our results we rely on a nonparametric Wilcoxon rank test (64-66). The difference between the bias of the high-earning bankers and that of the middle-earning bankers is significant, with a z-statistic of 3.33 and a p-value of 0.001. The difference between high-earning and low-earning bankers is significant, with a z-statistic of 2.37 and a p-value of 0.019. The difference between low and mid earning is not significant, with a z-statistic of 1.43 and a p-value of 0.15. This lack of significance is consistent with the skewness of the earning distribution mentioned above. Some stars receive very high compensation, ensuring exceptional performance, while the bulk of the bankers have relatively similar compensation. Consequently, there is only a small difference between the monetary compensation of the low-earning category and that of the middle-earning category.

Table 5 reports the estimated probabilities of the performance of the bankers conditionally on their biases. We classified participants as having low (very high bias) bias if they were among the 35% (less than 15%) highest (lowest) paid of their occupational group. Conditional probabilities were estimated as empirical frequencies. Panel A reports the results obtained for the whole banker sample. Consistently with the hypothesis that the low-bias bankers outperform, the largest number in each column is on the first diagonal of bankers with the lowest bias are most likely to be low-bias; bankers with medium bias are most likely to be mid-bias; and bankers with the highest bias are most likely to be high-bias. To test the null hypothesis that performance is independent of bias, we perform a Chi square test. Since there are three rows and three columns in Table 5, the relevant statistic is a Chi Square with 4 degrees of freedom, for which the critical value at the 5% level is 9.49. The difference between the empirical contingency table and the theoretical one (computed under the null) was found to be 16.55. The null hypothesis that performance and bias are independent can thus be rejected at the 5% level.

21
Experience: Experience could influence the bias of the bankers and its impact on their performance. For example, Lut (2009) provides evidence on the consequences of career experience on market anomalies. To control for experience we split the sample of bankers in two groups. The first group includes the bankers with less than 10 years of experience ("junior") and the second group includes the bankers with 10 years of experience or more ("senior"). For junior, the median bias is 28 for bankers with low earnings, 45 for bankers with medium earnings and 68 for bankers with high earnings. For seniors, the median bias is 27 for bankers with low earnings, 48 for bankers with medium earnings and 73 for bankers with high earnings. Thus, the result that bankers with high performance are less biased obtains in both age groups. Our findings also suggest that experience does not reduce the bias.

Job: Some participants in the experiment were employed in research, others in sales, and yet others in trading (or sales and trading). As an additional robustness check, we analyzed the control settings for each of these three different occupational categories. Panel D of Table 5 split the estimation of conditional probabilities according to occupational categories. The results suggest the link between bias and performance is reasonably robust across occupational categories in the bank. Consistently with the hypothesis that the backlight bias reduces performance, bankers with the lowest bias are most likely to be high-commissions for research and for sales, bankers with medium bias are most likely to be mid-commissions, for sales and for trading. And bankers with the highest bias are more likely to be low-commissions for research and for trading. In addition, for research and for trading, no banker with high bias has high earnings.
Minimizers. Yet another robustness check is to examine whether our results still hold after one
controls for the provision of the information of the participants relative to the questions asked.
As discussed above, for each participant, we computed a minimizer ratio, decreasing with the
provision of the answers of the agent in the own answer treatment. Interestingly, the minimise
ratio varies somewhat with performance. Participants with high performance have an average
minimiser ratio of 0.98% while participants with medium performance have a minimiser ratio of
1.04%, and participants with low performance have a minimiser ratio of 1.50%. Yet, the correlation
between the minimiser ratio and the hindsight bias is not significant. The Wilcoxon rank-sum test
comparing the minimiser ratio of high and medium earnings in \( z = -0.44 \) with a \( p \)-value of 0.66.
The Wilcoxon rank-sum test comparing the minimiser ratio of high and low earnings is \( z = 1.28 \),
with a \( p \)-value of 0.20. And the Wilcoxon rank-sum test comparing the minimiser ratio of
medium and low earnings in \( z = 2.04 \), with a \( p \)-value of 0.04. Furthermore, as mentioned
above, the minimiser ratio is not significantly correlated with the hindsight bias.

To control for stable biases, we divided the population of players into three categories. The
first category includes the 29 minimisers with the lowest minimise ratios (their average minimise
ratio is 0.28%). The second category includes the 29 minimisers with intermediate minimise ratios
(their average minimise ratio is 0.59%). The third category includes the 29 minimisers with the
highest minimise ratios (their average minimise ratio is 1.79%). We also divided players in three
bias categories similarly to Table 5. Within each "minimum category", the probability to be in the high-earnings category condition on their bias. We found that, in the
the three "minimum categories", players with the lowest bias were the most likely to be high-earners.
For the 39% lowest bias minimisers, the probability of being in the high earnings category is 30%.
For the middle minimisers group, 26% for the high "minimum category" and 23% for the high
"minimum category". For the medium and high bias minimisers, the probability of being in the
high earnings category is between 20% and 33% across all "minimum categories".

21
Overconfidence: Other psychological biases than the hindsight bias could affect performance. To control for such possible effects, we use the two other psychometric variables we measure: the better than average score and the misattribution error. Table 6 presents their means and median for the three earnings groups. As can be seen in the table, there is no strong variation in biases from one earnings group to the other. The median misattribution score is .376 for low earnings bankers and .500 for high earnings bankers. That is, for low earnings bankers the true answer $$4$$ is outside the bounds given by the participants 72.6% of the time, while for the high earnings bankers it is for 69.1% of the time. The median better than average score is .417 for the low earnings bankers and .301 for the high earnings bankers. That is, low earnings bankers assessed that 30.7% of their colleagues were better than themselves, while high earnings bankers assessed that 39.4% of their colleagues were better than themselves. Such small (and non-significant) variations from one earnings group to the other contrast with the large variations observed for the median hindsight bias, which falls from .386 for the low earnings bankers to .375 for the high earnings bankers.

5 Conclusion

Do cognitive biases affect information processing and performance in financial markets? In this paper we addressed that question, focusing on the hindsight bias. Agents who exhibit this bias fail to remember how ignorant they were before observing outcomes and answers. We show that this hinders learning, and, in particular, lead agents to underestimate volatility. This results in inefficient portfolio choice, loss-making trades, and poor risk management. We rely on two experimental studies to test these claims.

In the first experimental study, we focus on the consequences of the hindsight bias on learning. Unfortunately, we did not collect data on this variable.
about volatility. We compare two treatments: one in which the bias is muted, and the other where it can manifest itself. Agents give lower volatility updates in the latter treatment than in the former, as implied by our model.

In the second experimental study, we test the hypothesis that the hindsight bias hurts financial performance. We collect psychometric and performance data about highly paid investment bankers. We find that they exhibit hindsight bias when asked questions about economic trends and financials and that experience does not reduce this bias. Most importantly, we find that bankers with low bias obtain significantly better performance.

Given that our results suggest the hindsight bias matters, it would be interesting to provide a deeper theoretical analysis of this bias. Could it emerge from cognitive limitations, such as limited memory? Can our model be extended to accommodate a more general "projection bias" phenomenon (see Loewenstein, O'Doherty and Rubia, 2000)? And how would biased agents strategically interact in a trading game? We leave these issues for further research.
References


comparisons, self-monitoring, and trading performance in an experimental 


Choice: Some Problems in Learning from Experience.” Academy of 
Management Journal, 30, 529-542.


Knowledge in Economic Settings: An Experimental Analysis.” 


Difference Analysis.” Journal of Personality, 50, 45-60.


Organizational Behavior and Human Decision Processes, 48, 71-100.

Unrealistic perceptions of control and trading performance.” Journal of Organizational and 
On-


Table 1: Hindsight bias and risk perception in Study 1

On December 6, 2005, participants were asked to predict prices one week later and give confidence intervals, from which we calculated implied estimates. On December 13, they performed the same task again. In the second treatment, they were told their previous estimates. In the second week they were asked to reestimate their estimates. The table reports mean across agents. *p-values are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Satis</th>
<th>Risk</th>
<th>Risk perception</th>
<th>Low</th>
<th>Conf.</th>
<th>Implied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial implied</td>
<td>1.42</td>
<td>1.47</td>
<td>1.48</td>
<td>1.77</td>
<td>1.72</td>
<td>1.48</td>
</tr>
<tr>
<td>Second treatment</td>
<td>1.40</td>
<td>1.35</td>
<td>1.39</td>
<td>1.88</td>
<td>1.99</td>
<td>1.40</td>
</tr>
<tr>
<td>Task 1</td>
<td>21.1</td>
<td>25</td>
<td>20.14</td>
<td>23</td>
<td>26.3</td>
<td>21.10</td>
</tr>
<tr>
<td>Task 2</td>
<td>27.1</td>
<td>8</td>
<td>20.17</td>
<td>25</td>
<td>28</td>
<td>27.10</td>
</tr>
<tr>
<td>Task 3</td>
<td></td>
<td></td>
<td></td>
<td>33.6</td>
<td>33.6</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Randomization in Study 2

<table>
<thead>
<tr>
<th>Location</th>
<th>Randomization</th>
<th>Number</th>
<th>Average Age</th>
<th>Average Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT TWIN</td>
<td>High</td>
<td>12</td>
<td>34.34</td>
<td>5350</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>17</td>
<td>34.37</td>
<td>4250</td>
</tr>
<tr>
<td>MT NON</td>
<td>High</td>
<td>14</td>
<td>34.46</td>
<td>5448</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>16</td>
<td>33.38</td>
<td>4349</td>
</tr>
</tbody>
</table>

20
Table 3. B datasets questions and results in Study 2.

The figures in the first column compare the answers of the panelists. The first column reports the true answer to the question, the second column reports the majority of the group who answered that question, the third column reports the guess of the other group about that answer. The last column reports the likelihood bias, measured using Converse et al. (1990) index.

<table>
<thead>
<tr>
<th>Panel A: Panelist questions</th>
<th>True answer</th>
<th>Other group answer</th>
<th>False group prediction (given true answer)</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer price change</td>
<td>2.5</td>
<td>1.25</td>
<td>228</td>
<td>1.29</td>
</tr>
<tr>
<td>Money stock market from</td>
<td>3.0</td>
<td>2.25</td>
<td>1.74</td>
<td>0.00</td>
</tr>
<tr>
<td>Change in stock prices</td>
<td>2.2</td>
<td>1.00</td>
<td>1.33</td>
<td>1.00</td>
</tr>
<tr>
<td>Number of households</td>
<td>300</td>
<td>300.25</td>
<td>100.24</td>
<td>0.00</td>
</tr>
<tr>
<td>% of monthly earnings</td>
<td>10.1</td>
<td>31.22</td>
<td>21.29</td>
<td>0.22</td>
</tr>
<tr>
<td>Change in stock prices</td>
<td>10.1</td>
<td>10.1</td>
<td>10.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Number of households</td>
<td>354</td>
<td>194.5</td>
<td>169.5</td>
<td>0.00</td>
</tr>
<tr>
<td>Net revenue minus</td>
<td>905</td>
<td>1014.5</td>
<td>905.5</td>
<td>40.45</td>
</tr>
<tr>
<td>Medicaid income and GDP</td>
<td>821</td>
<td>535</td>
<td>582</td>
<td>30.30</td>
</tr>
<tr>
<td>Oecd income and GDP</td>
<td>641</td>
<td>545</td>
<td>505</td>
<td>14.27</td>
</tr>
</tbody>
</table>
Table 1: London sample

<table>
<thead>
<tr>
<th>Question</th>
<th>True answer</th>
<th>Other group answer</th>
<th>Three-group prediction</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail Price Index August 1990 to 2000</td>
<td>1.27</td>
<td>0.27</td>
<td>0.81</td>
<td>0.46</td>
</tr>
<tr>
<td>Drop of Lumia mobile phone prices high to 0-5%</td>
<td>8.17</td>
<td>9.16</td>
<td>9.14</td>
<td>1.25</td>
</tr>
<tr>
<td>Change in price of oil from 2010 to 2016</td>
<td>6.28</td>
<td>3.88</td>
<td>6.48</td>
<td>0.21</td>
</tr>
<tr>
<td>Numbers at London in August 2001</td>
<td>1.00</td>
<td>1.003</td>
<td>0.428</td>
<td>0.42</td>
</tr>
<tr>
<td>%锨服 qualche earnings from asset management, 2011</td>
<td>1.01</td>
<td>2.88</td>
<td>1.58</td>
<td>1.22</td>
</tr>
<tr>
<td>Bank Stock price index 2/1/2002</td>
<td>1.004</td>
<td>1.425</td>
<td>1.222</td>
<td>0.28</td>
</tr>
<tr>
<td>Rate of profit before tax in CEE countries in 2000</td>
<td>38.44</td>
<td>39.24</td>
<td>38.88</td>
<td>0.26</td>
</tr>
<tr>
<td>Net return before tax of 2nd quarter 2001</td>
<td>1.06</td>
<td>1.36</td>
<td>1.17</td>
<td>0.06</td>
</tr>
<tr>
<td>Growth rate of stock in CEE countries, 2001</td>
<td>5.0</td>
<td>4.25</td>
<td>4.92</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 2: Hit/miss bias for the three earnings categories in Study 2

The bias is measured using Conover et al. (1980). Median, mean and standard dev. are computed.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pool</th>
<th>President</th>
<th>London</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>Low</td>
<td>Med</td>
<td>High</td>
</tr>
<tr>
<td>Median</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>StDev</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Min</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Max</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3: Probability of performance, conditional on hit/miss bias in Study 2

The hit/miss were split in three categories of equal size. The first one included the 385 bottom 2nd, while the third one included the 385 top 2nd. Empirical frequencies are used to estimate conditional probabilities. The probabilities in each column add up to 1.
### Panel A: All Incomes

<table>
<thead>
<tr>
<th>Low Incomes</th>
<th>Medium Incomes</th>
<th>High Incomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob(High Earnings)</td>
<td>.40</td>
<td>.42</td>
</tr>
<tr>
<td>Prob(Med Earnings)</td>
<td>.14</td>
<td>.20</td>
</tr>
<tr>
<td>Prob(Low Earnings)</td>
<td>.26</td>
<td>.20</td>
</tr>
</tbody>
</table>

### Panel B: Residuals

<table>
<thead>
<tr>
<th>Low Incomes</th>
<th>Medium Incomes</th>
<th>High Incomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob(High Earnings)</td>
<td>.27</td>
<td>.35</td>
</tr>
<tr>
<td>Prob(Med Earnings)</td>
<td>.14</td>
<td>.21</td>
</tr>
<tr>
<td>Prob(Low Earnings)</td>
<td>.20</td>
<td>.28</td>
</tr>
</tbody>
</table>

### Panel C: Sides

<table>
<thead>
<tr>
<th>Low Incomes</th>
<th>Medium Incomes</th>
<th>High Incomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob(High Earnings)</td>
<td>.27</td>
<td>.35</td>
</tr>
<tr>
<td>Prob(Med Earnings)</td>
<td>.14</td>
<td>.21</td>
</tr>
<tr>
<td>Prob(Low Earnings)</td>
<td>.20</td>
<td>.28</td>
</tr>
</tbody>
</table>

### Panel D: Trading and Sides Trading

<table>
<thead>
<tr>
<th>Low Incomes</th>
<th>Medium Incomes</th>
<th>High Incomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob(High Earnings)</td>
<td>.27</td>
<td>.35</td>
</tr>
<tr>
<td>Prob(Med Earnings)</td>
<td>.14</td>
<td>.21</td>
</tr>
<tr>
<td>Prob(Low Earnings)</td>
<td>.20</td>
<td>.28</td>
</tr>
</tbody>
</table>

Table 6: High/low trade, misclassification and better than average for the three earnings categories in Study 2

<table>
<thead>
<tr>
<th>Low Earnings</th>
<th>Medium Earnings</th>
<th>High Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.30</td>
<td>0.36</td>
</tr>
<tr>
<td>Median</td>
<td>0.0001</td>
<td>0.2300</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.2367</td>
<td>0.2294</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.0650</td>
<td>0.0605</td>
</tr>
</tbody>
</table>

32