WWP

Wolfsburg Working Papers

No. 16 - 3

On the Measurement of Overconfidence

An Experimental Study

Markus Spiwoks and Kilian Bizer, October 2016
On the Measurement of Overconfidence

An Experimental Study

Markus Spiwoks and Kilian Bizer

Abstract

This study addresses two methods commonly used to measure overconfidence and examines their reliability. It is shown that traditional methods of assessing subjective certainty concerning a given forecast can contribute to a systematic overestimation of overconfidence. It also becomes apparent that a common approach measuring relative overconfidence can cause a distortion of scientific results. This is particularly valid for study groups in which female and male participants are not equally represented.

Keywords: Behavioral Finance; Experiments; Overconfidence; Stock Price Forecasts; Gender Studies.

JEL classification: G02, G17, D81, D84, C99.
1. Introduction

Economic subjects tend to overestimate their own abilities. This behavior has repeatedly been proven by numerous studies (an overview on corresponding literature can be found in Thoma, 2016, pp. 221-222; Morales-Camargo et al., 2015, pp. 69-70; Brookins, Lucas and Ryvkin, 2014, pp. 1-2; Bar-Yosef and Venezia, 2014, pp. 79-80; Oberlechner and Osler, 2012, pp. 91-95; Eshraghi and Taffler, 2012, pp. 2-3; Deaves, Lüders and Schröder, 2010, pp. 402-403; Blavatskyy, 2009, pp. 39-41; Kogan, 2009, pp. 1892-1894; Dittrich, Güth and Maciejovsky, 2005, pp. 471-473; Barber and Odean, 2001, pp. 262-265). Finding a way of realistically measuring overconfidence has ever been a major concern in scientific research (cf. e.g. Olsson, 2014; Parker and Stone, 2014; Michailova and Katter, 2014; Huisman, van der Sar and Zwinkels, 2012; Merkle and Weber, 2011; Benoît and Dubra, 2011; Speirs-Bridge et al., 2010; Cesarini, Sandewall and Johannesson, 2006). The present experimental study is intended to be a further contribution to this ongoing debate. It focuses on the influence of possible responses on the absolute and relative overconfidence or participants. It is shown that the extent of overconfidence, if measured conventionally, is rather overestimated than underestimated. This paper therefore warns against overconfidence in measuring overconfidence.
2. Hypothesis and Experimental Design

The participants were supposed to forecast the development of prices for five stocks. These were (1) the US-American biotech company Gilead Sciences Inc., (2) the US-American social network Facebook Inc., (3) the Russian oil company Lukoil Neftyanaya Komp., (4) the German information-technology company Bechtle AG and (5) the Chinese high street bank Bank of China.

Each of the participants is provided with very short information about the companies as well as the current stock prices (closing prices of the previous day). They are supposed to estimate whether the stock prices (a) rise or (b) fall or hold steady until a due date that is set approximately six weeks in the future.

Furthermore, the participants have to predict the development of two fictional equity funds which solely invest in the five stocks analyzed before. They are informed about the structure of the funds: 12.5% (25%) of the fund “Worldwide ZZX-2” (“Global PPS-1”) are Gilead stocks, 12.5% (16%) Facebook stocks, 25% (17%) Lukoil stocks, 25% (25%) Bechtle stocks and 25% (17%) Bank of China stocks. For these funds, the participants were again asked to forecast if the prices would (a) increase or (b) drop or remain constant until the due date of the forecast.

The participants are then supposed to self-assess the accuracy of their own forecasts. The alternative responses are presented in steps of 10%. It is common to only offer the alternative responses from 50%-100% since the possible responses from 0% to 40% are meaningless. This strategy is based on the following assumption: there are only two possible future events. Either the prices increase (alternative A) or the prices drop or hold steady (alternative B). If an economic subjects chooses alternative A and is only 40% certain of their decision, it would make sense to choose alternative B which then had to come true with a subjective certainty of 60%.

These settings are quite easy to understand which is why the participants should be able to assess their accuracy of forecasting independently from the alternative responses from 50% to 100% or from 0% to 100%. Therefore, hypothesis 1 reads as follows: a scale of subjective certainty of forecasting to which pointless alternative responses have been added does not influence the way the participants assess the accuracy of their own forecasts.
After the first part of the experiment, the participants are asked to self-assess their ability of forecasting in comparison to the other participants. Half of them get to choose between the responses “above average” and “below average”. “Above average” (“below average”) means that more (less) correct forecasts were given than the average number of correct forecasts by all participants. The other half of participants get to choose from the alternative responses: “above average”, “average” and “below average”. The alternative response “average” is defined as follows: if they start with the average value of correct forecasts minus half a standard deviation and end with the average value of correct forecasts plus half a standard deviation, the results of the forecasts are average. Therefore, slightly more than a third of all forecasts should fall into this category.

As yet, there is no indication that the offer of two or three alternative responses has an impact on overconfidence. Therefore, hypothesis 2 reads as follows: The extent of relative overconfidence does not vary in dependence of the number of alternative responses offered (two or three alternatives).

The experiment was conducted in two parts to avoid the dependency of the results from a single situation on the capital market. The first part took place on 22, 23 and 24 April 2015. The participants forecasted the price development until 7 June 2015, which is a prognosis period of about six weeks. The second part of the inquiry was conducted on 27/28/29 May 2015. The participants forecasted the price development until 10 July 2015, which again is a prognosis period of about six weeks. 240 students of business administration of the Ostfalia University of Applied Sciences participated in the experiment. Those 30 students with the most exact forecasts were rewarded € 50 each. Among the participants who assessed their forecasts correctly, twelve were chosen by lot and rewarded with another € 50. The total sum of rewards was € 2,100 which equates to € 8.75 for each participant. The experiment took approximately 20 minutes to complete, wherefore the reward can be deemed rather attractive. All participants seemed motivated and eager to give the best prognoses possible and to provide a precise relative self-assessment. Since the data collection was conducted as a classroom experiment, the opportunity costs for the participants were very low which is why there was no show-up fee. The participation in the experiment was voluntary.

In both prognosis periods, the prices developed as follows (table 1). The stock price
of Gilead Sciences Inc. increased in the period from 22 April to 7 June 2015 from € 97.96 to € 102.33. The stock price of Facebook, however, dropped from € 77.86 to € 73.75.

Table 1: Price Development of the Analyzed Stocks and Funds in the Prognosis Periods

<table>
<thead>
<tr>
<th>Experiment Part I: 22/23/24 April 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price on</strong></td>
</tr>
<tr>
<td>22.04.15</td>
</tr>
<tr>
<td>Gilead Sciences Inc.</td>
</tr>
<tr>
<td>Facebook Inc.</td>
</tr>
<tr>
<td>Lukoil Neftyanaya</td>
</tr>
<tr>
<td>Bechtle AG</td>
</tr>
<tr>
<td>Bank of China 100s</td>
</tr>
<tr>
<td>Fund ZZX-2</td>
</tr>
<tr>
<td>Fund PPS-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment Part II: 27/28/29 May 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Price on</strong></td>
</tr>
<tr>
<td>27.05.15</td>
</tr>
<tr>
<td>Gilead Sciences Inc.</td>
</tr>
<tr>
<td>Facebook Inc.</td>
</tr>
<tr>
<td>Lukoil Neftyanaya</td>
</tr>
<tr>
<td>Bechtle AG</td>
</tr>
<tr>
<td>Bank of China 100s</td>
</tr>
<tr>
<td>Fund ZZX-2</td>
</tr>
<tr>
<td>Fund PPS-1</td>
</tr>
</tbody>
</table>

real course = price development from the time when the prognosis was given to the end of the prognosis period; ↓ = price has dropped during the prognosis period; ↑ = price has risen during the prognosis period; 100s = block containing 100 stocks.
3. Results

It is shown that hypothesis 1 needs to be rejected. The alternative responses given (50%-100% or 0%-100%) immensely influenced the self-evaluation of the participants’ subjective certainty (figure 1). 15.86% of participants give useless estimates when offered the alternative responses 0%-100%. There are only two possible future events (increasing price or decreasing/steady price), wherefore decisions that are made with a subjective certainty of less than 50% are useless. In such cases, the participants should have chosen the other alternative at a time because the subjective certainty would then have been more than 50%.

It can be assumed that a few participants did not consider some of the presented alternative responses to be futile. Instead, they classed themselves according to their subjective certainty regarding the dependability of their forecasts. Some of those who felt rather uncertain, made senseless decisions and chose the alternative responses from 0% to 40%. In both versions of the game (alternative responses 0%-100% and 50%-100%), we can observe that participants tend to settle slightly above average. In the version with 0%-100%, the alternative responses 50%, 60%, 70% and 80% are chosen most often (figure 1, gray bars). The alternative responses 50%, 60% and 70% are even chosen more often than in the version with 50%-100%. In that version, many participants settle on average. The alternative responses 70% and 80% were chosen most often (figure 1, black bars).

**Figure 1:** Frequencies of Chosen Alternative Responses Relating to the Subjective Certainty of Price Forecasts
The discrepancy between the chosen responses in both versions of the game indicate that the participants do not really account for the subjective probabilities with which their forecasts will come to pass. They spontaneously class themselves in the given array instead. This means that, with the traditionally offered alternative responses (50%-100%), at least some participants abandoned themselves to the impulse to class themselves as average or slightly above average. This behavior would lead, however, to a systematic overestimation of overconfidence.

**Figure 2: Self-Assessment of Relative Forecast Success (All Participants)**

The analysis of the relative overconfidence produces a surprising result. The entire group of participants does not show any relative overconfidence, regardless of whether two alternative responses (game version 1, figure 2) or three alternative responses (game version 2, figure 3) are offered. In the first version 59 participants exactly classed themselves in each of the categories “above average” and “below average”.

In the second version of the game, 27 participants classed themselves as “above average”, 70 participants as “average” and 24 participants as “below average”. Hence, we cannot detect a collective overconfidence in either version of the game, which supports hypothesis 2.
These results are in stark contrast to those regarding the relative self-assessment in studies that address general abilities (from car driving to expressive powers; cf. e.g. Svenson, 1981) and are thereby rather surprising (similar results were obtained by Proeger and Meub, 2014; Clark and Friesen, 2009). Taking a closer look at the results, we can see that there is a clear discrepancy concerning the relative self-assessment of male and female participants.

While the male participants reveal a strong tendency towards overconfidence when giving their self-assessment in game version 1 (two alternative responses), the female participants show the exact opposite behavior. In game version 1 (two alternative responses), they show a clear tendency towards underconfidence (figure 4). By the way, the actual forecast success is the same for both male and female participants.
These results are similar to those for game version 2 (three alternative responses). The male participants show a strong tendency to be overconfident whereas the female participants clearly are prone to be underconfident (figure 5).

These results confirm the former scientific findings which, too, revealed men to be more overconfident than women (cf. e.g. Santos et al., 2010; Johnson et al., 2006; Biais et al., 2005; Pulford and Colman, 1997). However, the results contravene those by Sharma and Shakeel (2015) or by Kufepaksi (2011), who demonstrated a stronger tendency for overconfidence in women.

Figure 4: Self-assessment of Relative Forecast Success by Male Participants (Black Bars) and Female Participants (Gray Bars) in Game Version 1 with Two Alternative Responses

Figure 5: Self-Assessment of Relative Forecast Success by Male Participants (Black Bars) and Female Participants (Gray Bars) in Game Version 2 with Three Alternative Responses
When comparing both versions of the game (two or three alternative responses) it can be observed that game version two has a cushioning effect on the overconfidence in male and on the underconfidence in female participants (tables 2 and 3). In game version 1, the male participants show a superfluity of overconfidence of well 37 percentage points (68.52% in the category “above average” and 31.48% in the category “below average”). In game version 2, however, this superfluity is reduced to 21 percentage points (32.39% in the category “above average” and 11.27% in the category “below average”). The male participants obviously find it hard to classify they relative success as below average which is why the classification “above average” are predominant in game version 1 (two alternative responses). Are they, however, offered a third response “average” (game version 2), their superfluity of overconfidence significantly recedes (from about 37 to 21 percentage points).

The situation is different for the female participants who clearly find it difficult to class themselves as above average in comparison to the other participants (tables 2 and 3). This is the reason why, with well about 31 percentage points, they show a strong superfluity of underconfidence (34.38% in the category “above average” and 65.63% in the category “below average”). If, however, they are offered the category “average” in game version 2, the superfluity of underconfidence is reduced to 24 percentage points (8.0% in the category “above average” and 32.0% in the category “below average”). Offering the alternative response “average” therefore reduces the superfluity of overconfidence in female participants from 31 to 24 percentage points.

Table 2: Relative Self-Assessment of Participants in Game Version 2 (Two Alternative Responses)

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
</tr>
<tr>
<td>Above average</td>
<td>37</td>
<td>68.52%</td>
<td>22</td>
</tr>
<tr>
<td>Below average</td>
<td>17</td>
<td>31.48%</td>
<td>42</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>100.00%</td>
<td>64</td>
</tr>
</tbody>
</table>
Table 3: Relative Self-Assessment of Participants in Game Version 2 (Three Alternative Responses)

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th></th>
<th>Female</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
<td>Percentage</td>
<td>Number</td>
<td>Percentage</td>
</tr>
<tr>
<td>Above average</td>
<td>23</td>
<td>32.39%</td>
<td>4</td>
<td>8.00%</td>
<td>27</td>
<td>22.31%</td>
</tr>
<tr>
<td>Average</td>
<td>40</td>
<td>56.34%</td>
<td>30</td>
<td>60.00%</td>
<td>70</td>
<td>57.85%</td>
</tr>
<tr>
<td>Below average</td>
<td>8</td>
<td>11.27%</td>
<td>16</td>
<td>32.00%</td>
<td>24</td>
<td>19.83%</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>100.00%</td>
<td>50</td>
<td>100.00%</td>
<td>121</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

If male and female participants are not equally represented, it can be helpful to offer the third alternative response (“average”) in addition to the usual two alternatives (“above average” and “below average”) in order to avoid a systematic falsification of results when analyzing any relative overconfidence.
4. Summary

The present study analyses two traditional methods of measuring overconfidence and assesses their reliability. The first method analyzed covers absolute overconfidence. The participants were given the task to forecast the price development of five real stocks and two fictional funds based on these stocks. Since the prognosis period is about six weeks long, the forecasts are actually given ex ante. The participants are to estimate if the prices (a) increase or (b) drop or hold steady. They also have to state how confident they are regarding their own forecasts. Using a between-subjects approach, we analyze whether the addition of senseless answers (0%-40%) affects their self-assessment. It is shown that at least some participants do not carefully consider their subjective certainty with which their forecasts are going to come to pass. Instead, they spontaneously class themselves in the given array. This also means that the usual offer of alternative responses (50%-100%) makes some participants follow the impulse to class themselves as average or slightly above average. This, however, would lead to a systematic overestimation of overconfidence.

Moreover, this study addresses a method to measure relative overconfidence. At the end of the experiment the participants are asked to evaluate their own prognoses in comparison to the other participants’ forecasts. Two game versions are compared by a between-subjects approach. One half of the participants had to choose between the two usual alternative responses “above average” and “below average”. The other half was offered the third alternative response “average”. It was observed that the entire group of participants exhibited no overconfidence. For both the first and the second version of the game, the responses of the self-assessment were equally distributed; almost the same number of participants responded with “above average” and with “below average”.

We can observe a significant discrepancy between male and female participants. While men are clearly prone to collective overconfidence, women show a strong tendency towards underconfidence. Comparing the two game versions leads to the assumption that the male participants found it particularly difficult to class their own performance as below average. The female participants, however, had issues with classing their forecasts as above average. Introducing the third alternative response significantly reduced the superfluity of overconfidence in men and the underconfidence in women. Hence, it seems useful to always consider the third
alternative response “average”, at least when one of the sexes is overrepresented in a study group. By doing so, the actual extent of overconfidence and underconfidence can be assessed without the risk of overestimating it.

Overall, this study provides an indication that the traditional methods of measuring overconfidence that were analyzed here can well lead to a misjudgment of actual overconfidence.
5. Bibliography


Appendix: Instructions

Game Version 1

Your task is to forecast stock prices and prices of stocks in a fund.

A reward of € 50.00 is paid to the five participants who give the best forecasts in today’s inquiry.

GILEAD SCIENCES INC. Current price: € 97.96

Gilead Sciences Inc. is an independent company, operating globally in the biotech industry. They focus on developing therapeutic solutions for treating fatal infectious diseases.

Please tick the box.

☐ The stock price will increase until 7 June 2015.

☐ The stock price will decrease or hold steady until 7 June 2015.

How certain are you regarding your estimate? How probable do you believe your forecast to be? Please tick the box.

☐ 50%    ☐ 60%    ☐ 70%    ☐ 80%    ☐ 90%    ☐ 100%

Now we ask you to evaluate yourself.

Two of the participants in today’s inquiry who assess themselves correctly have the chance of winning € 50.

Please tick the box.

In comparison to the other participants in today’s inquiry, I believe to have forecast the stock prices and funds

☐ above average

☐ below average
Game Version 2
Your task is to forecast stock prices and prices of stocks in a fund.

A reward of € 50.00 is paid to the five participants who give the best forecasts in today’s inquiry.

Facebook INC. Current price: € 73.11
Facebook is a product of the company by the same name and, at the present time, the biggest social network in the world.

Please tick the box.

O  The stock price will increase until 10 July 2015.
O  The stock price will decrease or hold steady until 10 July 2015.

How certain are you regarding your estimate? How probable do you believe your forecast to be? Please tick the box.

O O O O O O O O O O
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Now we ask you to evaluate yourself.

Two of the participants in today’s inquiry who assess themselves correctly have the chance of winning € 50.

Please tick the box.

In comparison to the other participants in today’s inquiry, I believe to have forecast the stock prices and funds

O  above average
O  averagely
O  below average