Ostfalia
Hochschule für angewandte
Wissenschaften

Wolfsburg

Fakultät Wirtschaft

WWP

Wolfsburg Working Papers

No. 21-01

Reducing Algorithm Aversion Through Experience

Ibrahim Filiz, Jan René Judek, Marco Lorenz and Markus Spiwoks, January 2021

Reducing Algorithm Aversion Through Experience

Ibrahim Filiz, Jan René Judek, Marco Lorenz and Markus Spiwoks

Keywords: Algorithm aversion, overconfidence, operating experience, stock market forecasting, behavioral finance, experiments.

JEL classification: D83, D84, D91, G17, G41.

Abstract: In the context of an experiment, we examine the persistence of aversion towards algorithms in relation to learning processes. The subjects of the experiment are asked to make one share price forecast (rising or falling) in each of 40 rounds. A forecasting computer (algorithm) is available to them which has a success rate of 70%. Intuitive forecasts made by the subjects usually lead to a significantly poorer success rate. Feedback provided after each round of forecasts and a clear financial incentive lead to the subjects becoming better able to estimate their own forecasting abilities. At the same time, their aversion to algorithms also decreases significantly.

Ibrahim Filiz, Ostfalia University of Applied Sciences, Faculty of Business, Siegfried-Ehlers-Str**.** 1, D-38440 Wolfsburg, Germany, Tel.: +49 160 3344 078, E-Mail: ibrahim.filiz@ostfalia.de

Jan René Judek, Ostfalia University of Applied Sciences, Faculty of Business, Siegfried-Ehlers-Str. 1, D-38440 Wolfsburg, Germany, Tel.: +49 5361 892 225 420, E-Mail: ja.judek@ostfalia.de

Marco Lorenz, Georg August University Göttingen, Faculty of Economic Sciences, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany, Tel.: +49 1522 6672 503; E-Mail: marco.lorenz@stud.uni-goettingen.de

Markus Spiwoks, Ostfalia University of Applied Sciences, Faculty of Business, Siegfried-Ehlers-Str. 1, D-38440 Wolfsburg, Germany, Tel.: +49 5361 892 225 100, E-Mail: m.spiwoks@ostfalia.de

1. Introduction

Bank customers are becoming increasingly aware of charges, which is creating considerable cost pressure for banks. Particularly in the high-cost field of asset management, banks are endeavoring to reduce their personnel costs in relation to the provision of services to customers with low to medium amounts of assets. The substantial progress made in the field of artificial intelligence is increasingly leading banks to offer the services of so-called robo advisors which can provide customers with largely automated asset management (see, for example, Rühr et al., 2019; Jung et al., 2018; Singh & Kaur, 2017). There are some typical errors which are frequently made by professional investors as well as amateurs. For example, securities portfolios are often under-diversified (see, for example Dimmock et., 2016; Anderson 2013; Hibbert, Lawrence & Prakash, 2012: Goetzmann & Kumar, 2008), or portfolios are restructured too frequently (see, for example, Barber & Odean, 2001; Barber & Odean, 2000). Many stock market players tend to see patterns in the trends of prices on the capital markets when in reality there are none (see, for example Zielonka, 2004; Wärneryd, 2001; Gilovich, Vallone & Tversky, 1985; Roberts, 1959). In this way, their gut feeling often entices them into making suboptimal investment decisions (see, for example, Frydman & Camerer, 2016; Kudryavtsev, Cohen & Hon-Snir, 2013). Problematic behavioral tendencies of this kind can easily be avoided with a suitably programmed robo advisor. An offer of reliable and cheap asset management which also has a favorable risk-return profile can thus be made to clients (see, for example, Rossi & Utkus, 2020; Bhatia, Chandani & Chhateja, 2020; D'Acunto, Prabhala & Rossi, 2019; Beketov, Lehmann & Wittke, 2018; Uhl & Rohner, 2018).

However, many people have reservations about automated processes. This frequently also applies even when it is clearly recognizable that an algorithm (such as that in a robo advisor) achieves better results than when an expert has taken on this task. This phenomenon is referred to as algorithm aversion (see, for example Erlei et al., 2020; Ku, 2020; Köbis & Mossink, 2020; Castelo, Bos & Lehmann, 2019; Dietvorst, Simmons & Massey, 2018; Prahl & Van Swol, 2017; Dietvorst, Simmons & Massey, 2015). This problem also occurs when subjects have to decide whether they trust themselves or an algorithm more (see, for example Efendić, Van de Calseyde & Evans, 2020; Rühr et al., 2019; Dietvorst, Simmons & Massey, 2018; Dietvorst, Simmons & Massey, 2015). Even when there are clear indications that it is hardly possible to make better decisions than the algorithm over the longer term, many subjects still tend to trust themselves more.

It seems reasonable to suppose that overestimation of one's own abilities plays a significant role here. Algorithm aversion and overconfidence are thus presumably closely related phenomena. However, there is an opportunity here. Proeger & Meub (2014) show that financial incentives, repeated feedback and the gradual development of subjects' experience can help them to learn to assess their capabilities better. A learning process can thus lead to a reduction of overconfidence.

It thus seems feasible that algorithm aversion can also be decreased notably when decision-making situations repeat themselves, clear feedback is provided, and there are financial incentives. It is precisely this which is examined in this study on the basis of an experiment using repeated share price forecasts.

2. Experimental design and hypotheses

The subjects have the task of forecasting the price of a stock in 40 periods. However, they do not have to predict the exact price, only whether it will rise or fall. The price is always moving up or down, so there is never an unchanged price. Either the price rises or it falls. The price of the share is essentially determined by four fundamental influencing factors (A, B, C and D). However, these fundamental influencing factors are supplemented by a random influence ε (cf. Filiz, Nahmer & Spiwoks, 2019; Meub et al., 2015; Becker, Leitner & Leopold-Wildburger, 2009).

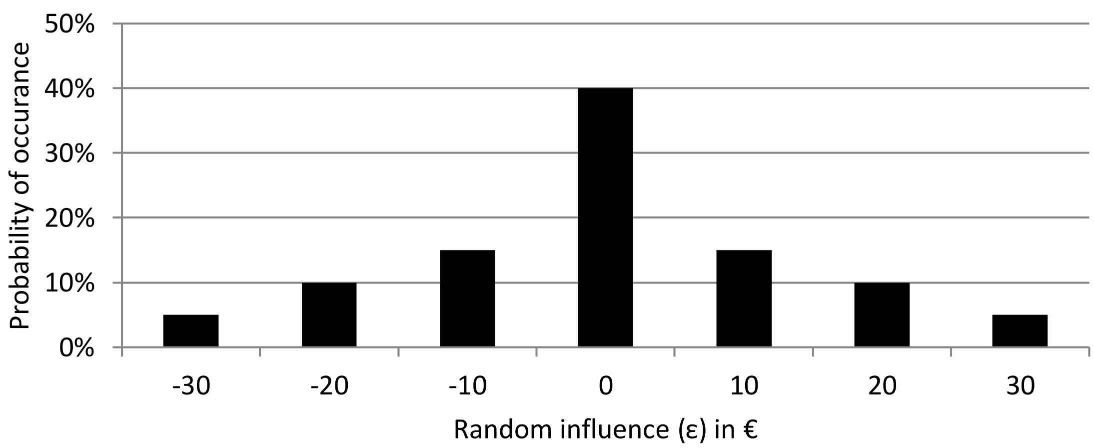
The price (K) of the share comes about as follows:

$K_t = 32\ A_t + 1\ B_t - 18\ C_t + 44\ D_t + \varepsilon_t$

The fundamental influencing factors (A, B, C and D) would, without the random influence ε_t, lead to a change in the price of between €0 and €10. If the fundamental influencing factors develop favorably overall, without the random influence ε_t there would always be a price increase between €0 and €10. This means that: €0 < (ΔK_t - ε_t) < +€10. However, if the fundamental influencing factors develop unfavorably overall, without the random influence ε_t there would always be a fall in the price of between €0 and €10. This means that: €0 > (ΔK_t - ε_t) > –€10.

The random influence ε_t has an expected value of 0 and exhibits the following distribution: with a probability of 40%, the random influence ε_t will not influence the price. With a probability of 15% each, the random influence ε_t will change the price by +€10 or by –€10. With a probability of 10% each, the random influence ε_t will change the price by +€20 or by –€20. And with a probability of 5% each, the random influence ε_t will change the price by +€30 or by –€30 (Figure 1).

Figure 1: Distribution of probability of the random influence ε_t



The fundamental influencing factors are announced to the subjects before each prediction round. In each round they have the opportunity to either make their own assessment (price rises or falls) or to delegate the decision to a forecasting computer (algorithm). In 70% of cases, the forecasting computer estimates the trend of the future share price correctly.

In other words, the algorithm merely exploits the available information about the fundamental influencing factors and the random influence ε_t in an optimal way. As the expected value of the random influence ε_t is zero, the algorithm calculates as follows: $K_t = 32\ A_t + 1\ B_t - 18\ C_t + 44\ D_t + 0$. Then it compares K_t with K_{t-1}. If $K_t > K_{t-1}$, the algorithm forecasts a rising trend. If $K_t < K_{t-1}$, the algorithm predicts a downward trend.

If the fundamental data suggests a rising trend (+€10 > ΔK_t > €0), this remains true in 70% of cases, also after the random influence ε_t is taken into consideration. A downward trend rather than a rising trend only transpires if the following random influences occur: ε_t = –€10 (15% probability) or ε_t = –€20 (10% probability) or ε_t = –€30 (5% probability). If the fundamental data suggests a downward trend (–€10 < ΔK_t < €0), this remains true in 70% of cases, also after the random influence ε_t is taken into consideration. An upward trend rather than a downward one only transpires when the random influences ε_t = +€10 (15% probability) or ε_t = +€20 € (10% probability) or ε_t = +€30 (5% probability) occur.

The algorithm thus uses the existing information optimally, but its forecasts are by no means perfect. It is only right in 70% of cases. The phenomenon of algorithm aversion appears particularly in the case of algorithms which obviously do not function perfectly (see, for example, Dietvorst, Simmons & Massey, 2015).

The subjects are given an insight into 40 periods of historical prices of stock Z before they have to make their first decision (see Appendix 3). In these 40 periods of price history, the price has risen exactly 20 times and has fallen exactly 20 times. This pattern remains in the subsequent 40 periods too: the price rises 20 times and falls 20 times. The subjects are not explicitly informed about this. However, by looking at the price history they can obtain an impression of how the share price has risen just as frequently as it has fallen.

The subjects are aware of the mechanism behind how the price is formed ($K_t = 32\ A_t + 1\ B_t - 18\ C_t + 44\ D_t + \varepsilon_t$) and about the probability distribution of ε_t. In addition, the subjects are made expressly aware of the fact that the forecasting computer (algorithm) makes a correct prediction in 70% of cases. Test questions are used to ensure that the subjects have understood this point of departure (see Appendix 2).

For a total of 40 times the subjects now have the choice whether to make their own forecast or to trust the algorithm. For every correct forecast they make or which they let the algorithm make for them (price rises or falls), the subjects receive a payment of 50 cents. They receive no payment if they or the algorithm make an incorrect forecast.

As the sequence of rising and falling price trends has no pattern which would enable a rational forecast (see Appendix 4), the subjects have a choice between three strategies, although they in no way have to stick to just one of them. In each round of forecasts they have a free choice as to how they act. It is only in this way that we can observe possible learning effects. These three strategies are basically as follows:

1. The subjects try to guess the trend of the price intuitively. In this case they would guess correctly in around 50% of cases. The expected value of their payment in this case is €10.
2. The subjects use all of the information available to them and make forecasts in the same way as the algorithm would. To support them in this strategy they are given a pocket calculator, a

pen and paper. In this case they will choose correctly in around 70% of cases. The expected value of their payment is €14.

3. They delegate the forecasting to the algorithm. In this case they will make a correct forecast in around 70% of cases. The expected value of their payment is €14.

Subjects who act rationally and maximize their utility (*homo oeconomicus*) would have to choose the third strategy. The first strategy leads to a noticeable reduction in the expected value of their payment. The second strategy does not lead to a higher expected value of the payment than the third strategy, but due to the considerable calculation work required (an overall total of 160 multiplications with 320 factors plus 40 additions of 160 summands) it is prone to errors and arduous. A rational subject will therefore undoubtedly choose the third strategy.

However, it is well-known from earlier studies that in many subjects, looking at the price history of a stock triggers a strong feeling of intuition about its possible future trend (see, for example, Zielonka, 2004; Wärneryd, 2001; Roberts, 1959). We therefore presume that by no means all subjects will stay with the third strategy from the first round of the game to the last.

Hypothesis 1 is therefore: Some subjects will – at least sometimes – not choose the third strategy (delegation of forecasting to the algorithm).

Null hypothesis 1 is therefore: All subjects will choose the third strategy (delegation of forecasting to the algorithm) in all forty rounds of the game.

We presume that algorithm aversion and overconfidence are similar behavioral anomalies. The subjects will therefore frequently follow their own intuition instead of the algorithm (first strategy) because they overestimate their own forecasting ability. If one takes into account the results of the research by Proeger and Meub (2014), it can be presumed that the subjects will gradually learn to assess their own forecasting ability more realistically, because after each round of forecasting they are informed about how the price has changed (rising or falling), how successful they have been with their decisions (the current amount of their payment), and how successful they would have been if they had always delegated the forecasts to the algorithm (see Appendix 3).

Hypothesis 2 is therefore: In the last 5 (10/15/20) rounds of forecasting, the subjects will trust the algorithm significantly more often than they did in the first 5 (10/15/20) rounds.

Null hypothesis 2 is therefore: In the last 5 (10/15/20) rounds of forecasting, the subjects will not choose the algorithm significantly more often than they did in the first 5 (10/15/20) rounds.

3. Results

The experiment is carried out between 2-14 November 2020 in the Ostfalia Laboratory of Experimental Economic Research (OLEW) of Ostfalia University of Applied Sciences in Wolfsburg. Overall, 143 subjects take part in the experiment. The subjects are students of Ostfalia University of Applied Sciences in Wolfsburg. 65 subjects (45.5%) study at the Faculty of Business, 60 subjects (42%) at the Faculty of Automotive Engineering, and 18 subjects (12.6%) at the Faculty of Public Health Services. 91 subjects (63.6%) are male, 50 subjects (35%) are female and 2 subjects (1.4%) assign themselves to the category of third gender. The youngest subject is 18. The oldest subject is 35. The average age of the subjects is 23.5 years.

The experiment is programmed with z-Tree (cf. Fischbacher, 2007). In the Ostfalia Laboratory for Experimental Economic Research (OLEW), there are twelve computer workplaces. However, only a maximum of four are used per session. This ensures that a considerable distance can be maintained between the subjects. This is necessary due to the Covid-19 pandemic so that there is no danger to the health of the subjects. The workplaces in the laboratory are also equipped with divider panels, which makes it possible to completely separate the subjects from each other. The experiments are constantly monitored by the experimenter so that communication between the subjects and the use of prohibited aids (such as smartphones) can be ruled out. Overall a total of 42 sessions are carried out. A session lasts an average of 45 minutes.

Is algorithm aversion exhibited in this experiment, or do all of the subjects consistently select the algorithm? A subject who is fully informed and always looking for their maximum utility (homo oeconomicus) would have to trust the algorithm in each of the 40 rounds of forecasting. This strategy leads to the maximum possible expected value in terms of the payment.

Overall, 143 subjects make 40 decisions each. This makes a total of 5,720 decisions. Of these, only 2,624 decisions (45.9%) are made in favor of the algorithm. In 3,096 decisions (54.1%), the subjects do not trust the algorithm. A clear majority of decisions are thus characterized by reservations in relation to the algorithm (Table 1).

Table 1: Decisions for and against the algorithm

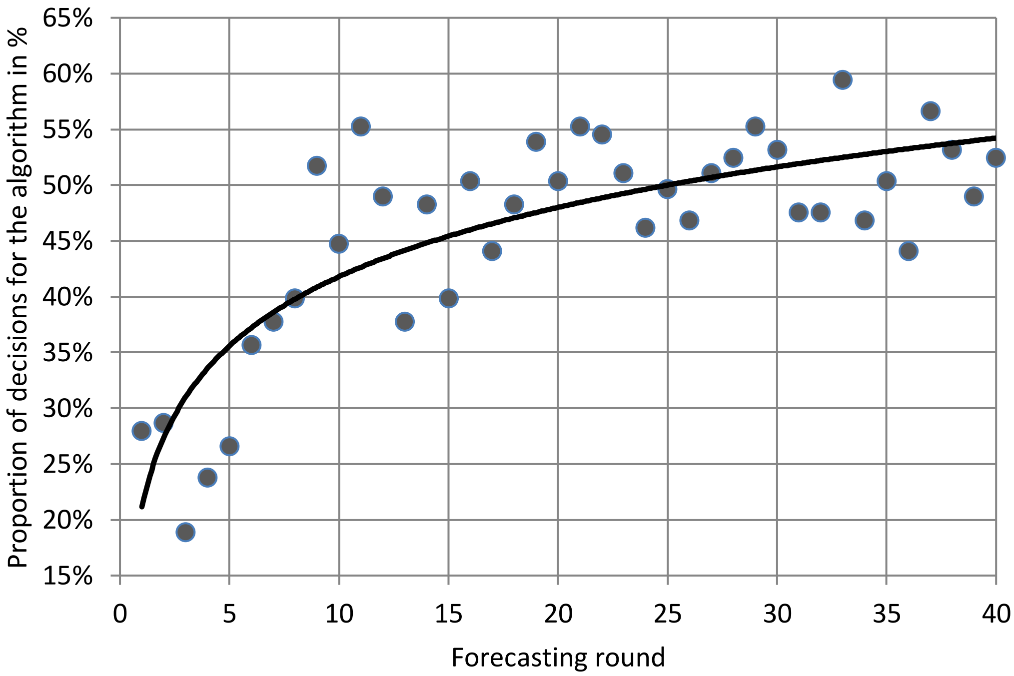
Decisions in favor of the algorithm		Decisions against the algorithm	
Number	%	Number	%
2,624	45.9%	3,096	54.1%

The t-test makes it clear that null hypothesis 1 has to be clearly rejected. The p-value of 0.000 underlines the clarity of the results: the presumption is thus confirmed that a considerable amount of algorithm aversion would be revealed and that the subjects by no means always make rational decisions which maximize their utility.

Only very few subjects consistently pursue the strategy in which the fundamental data is used in order to determine the expected value of the next share price and to make a comparison with the last actual price. We can only observe this behavior in five subjects (3.5%). Their decisions against the algorithm must, however, be fully attributed to the phenomenon of algorithm aversion, because for a fully-informed subject who wishes to maximize his or her utility, it is clearly recognizable that this strategy does not lead to a higher expected value of the payment. At the same time, it must be feared that errors can creep in given the multitude of calculations required (in 40 rounds a total of 160 multiplications with 320 factors and then 40 additions of 160 summands). That is why this tiresome mathematical recapitulation of the algorithm also reveals objectively unjustified reservations in relation to its reliability (Table 1).

Of particular interest is now whether the aversion to algorithms declines over time. Many subjects begin the experiment with an unjustified confidence that they can forecast the development of the share price (rising or falling) better than the algorithm. However, the sequence of rising and falling share prices is a random process with a probability of occurrence of 50% each for a rise or a fall of the price (see Appendix 4). No information about the next movements of the stock can be derived from the price history. In this respect, intuitive decisions lead to a significant reduction in the expected payment in the medium to long-term.

Figure 2: Proportions of decisions in favor of the algorithm in % according to forecasting rounds



After each round of forecasts, the subjects are informed about the success of the algorithm and if applicable about the success of their own diverging forecast. As time passes, it thus becomes increasingly clear to the subjects that trusting their own intuition and not the algorithm is a sub-optimal strategy. As the experiment proceeds, part of the subjects gives up their reservations in relation to the algorithm (Figure 2 and Table 2). If one inserts a logarithmic regression line (Figure 2), the characteristics of a typical learning curve (see, for example Anzanello & Fogliatto, 2011; Wright, 1936) with declining learning progress can be recognized.

Table 2: Decisions for and against the algorithm according to forecasting rounds

Forecasting round	Decisions for the algorithm		Decisions against the algorithm	
	Number	Percent	Number	Percent
1	40	27.97%	103	72.03%
2	41	28.67%	102	71.33%
3	27	18.88%	116	81.12%
4	34	23.78%	109	76.22%
5	38	26.57%	105	73.43%
6	51	35.66%	92	64.34%
7	54	37.76%	89	62.24%
8	57	39.86%	86	60.14%
9	74	51.75%	69	48.25%
10	64	44.76%	79	55.24%
11	79	55.24%	64	44.76%
12	70	48.95%	73	51.05%
13	54	37.76%	89	62.24%
14	69	48.25%	74	51.75%
15	57	39.86%	86	60.14%
16	72	50.35%	71	49.65%
17	63	44.06%	80	55.94%
18	69	48.25%	74	51.75%
19	77	53.85%	66	46.15%
20	72	50.35%	71	49.65%
21	79	55.24%	64	44.76%
22	78	54.55%	65	45.45%
23	73	51.05%	70	48.95%
24	66	46.15%	77	53.85%
25	71	49.65%	72	50.35%
26	67	46.85%	76	53.15%
27	73	51.05%	70	48.95%
28	75	52.45%	68	47.55%
29	79	55.24%	64	44.76%
30	76	53.15%	67	46.85%
31	68	47.55%	75	52.45%
32	68	47.55%	75	52.45%
33	85	59.44%	58	40.56%
34	67	46.85%	76	53.15%
35	72	50.35%	71	49.65%
36	63	44.06%	80	55.94%
37	81	56.64%	62	43.36%
38	76	53.15%	67	46.85%
39	70	48.95%	73	51.05%
40	75	52.45%	68	47.55%

It can be seen that the percentage of decisions for the algorithm is initially quite low. On average in the first five rounds of forecasts, only around a quarter of the decisions of the subjects (25.2%) are for the algorithm, but then a swift learning process begins. Many subjects recognize that their intuition is not sufficiently reliable. On average in rounds 6-10 the percentage of decisions in favor of the algorithm already rises to 42%. On average in rounds 11-15 the percentage of decisions in favor of the algorithm continues to rise to 46%.

The learning process and the gradual fading away of algorithm aversion take place above all in the first 20 rounds of forecasting (Figure 3). In the final 20 rounds of forecasting, however, there is no longer a significant reduction of algorithm aversion (Figure 4). In the first five rounds of forecasting the algorithm is chosen 180 times (25.2%) and in the last five rounds 365 times (51.1%). In the first 10 rounds of forecasting the algorithm is chosen 480 times (33.6%) and in the last 10 rounds 725 times (50.7%). In the first 15 rounds of forecasting the algorithm is chosen 809 times (37.7%) and in the last 15 rounds 1,095 times (51.1%). In the first 20 rounds of forecasting the algorithm is chosen 1,162 times (40.6%) and in the last 20 rounds 1,462 times (51.1%).

Figure 3: Percentage of decisions in favor of the algorithm in the first 5, 10, 15 and 20 rounds of forecasting

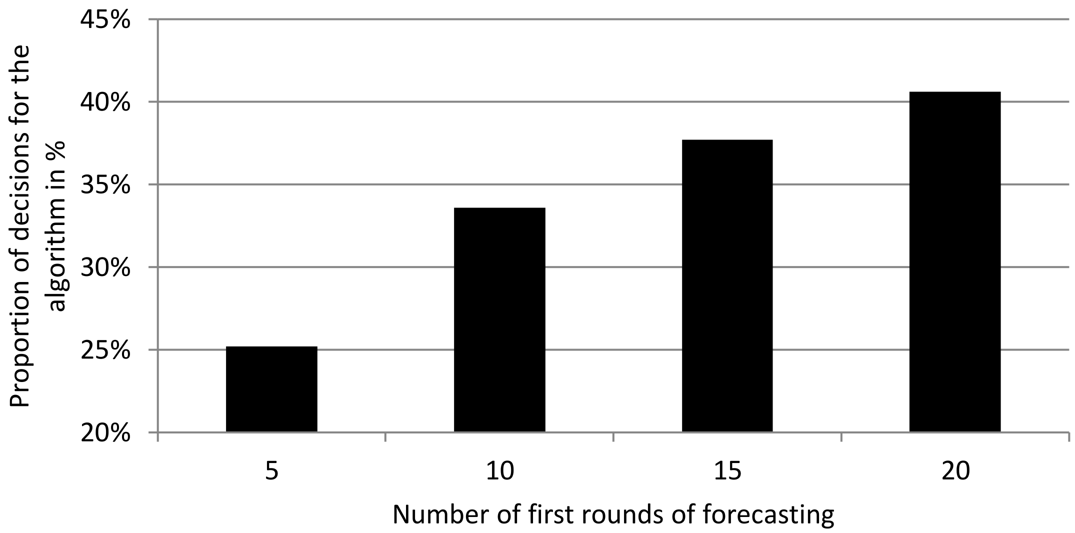
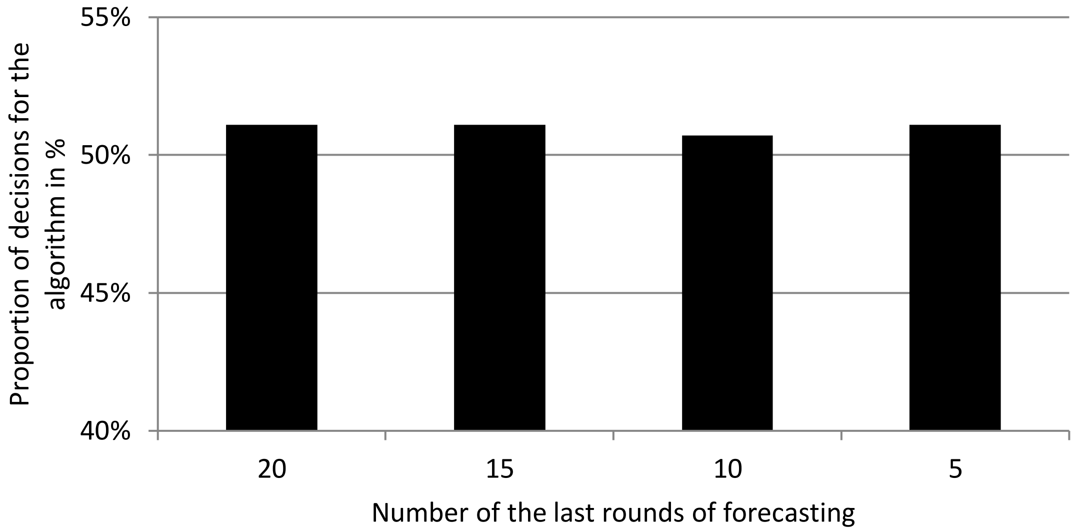


Figure 4: Percentage of decisions in favor of the algorithm in the last 20, 15, 10 and 5 rounds of forecasting



Hypothesis 2 can be checked with the help of a regression analysis. When carrying out a linear regression ($y_t = \beta_0 + \beta_1 \cdot x_t + u_t$) as well as when carrying out a logarithmic regression ($y_t = \beta_0 + \beta_1 \cdot \ln(x_t) + u_t$), it is clearly shown that algorithm aversion recedes over the course of 40 rounds of forecasting. The proportion of decisions in favor of the algorithm thus rises significantly. The p-values of the t-tests are unequivocal (Table 3). It is therefore clear that null hypothesis 2 has to be rejected. Algorithm aversion is significantly reduced during a learning process with a declining course.

Table 3: Regression analysis of the increase in decisions in favor of the algorithm

Regression	Regression equation	β_1	t-value	p-value
Linear	$y_t = \beta_0 + \beta_1 \cdot x_t + u_t$	+0,57	5,92	0,000***
Logarithmic	$y_t = \beta_0 + \beta_1 \cdot \ln(x_t) + u_t$	+8,96	8,59	0,000***

*** = significant with an error probability of 1%, ** = significant with an error probability of 5%,
* = significant with an error probability of 10%.

Another procedure for examining the significance of the learning process is the Wilcoxon signed-rank test. With the aid of this test it can also be established whether the gradual increase in the number of decisions in favor of the algorithm is statistically significant (Table 4).

Here, we observe the number of subjects who follow the algorithm in the last 5 (10/15/20) rounds of forecasting more frequently (less frequently/unchanged) than in the first 5 (10/15/20) rounds of forecasting. It has hardly any influence on the results whether one compares the first 5 rounds of forecasting with the last 5 rounds of forecasting, or whether one compares the first 10 rounds of forecasting with the last 10 rounds of forecasting, or whether one compares the first 15 rounds of forecasting with the last 15 rounds of forecasting, or whether one compares the first 20 rounds of forecasting with the last 20 rounds of forecasting. In all four cases, it can be seen that a learning process sets in over the course of the 40 rounds of forecasting. The subjects learn to assess their forecasting abilities more realistically. Algorithm aversion declines notably. In the Wilcoxon signed-rank test, the results prove to be highly significant (Table 4). Null hypothesis 2 clearly has to be rejected. Experience with the advantages of algorithms can thus certainly lead to a reduction of algorithm aversion.

Table 4: Decision-making behavior in the first and last rounds of forecasting

Number (x) of forecasting rounds considered (first and last)	Subjects with fewer decisions in favor of the algorithm in the first x forecasting rounds than in the last x forecasting rounds	Subjects with more decisions in favor of the algorithm in the first x forecasting rounds than in the last x forecasting rounds	Subjects with the same number of decisions in favor of the algorithm in the first x forecasting rounds as in the last x forecasting rounds	Σ	p-value Wilcoxon signed-rank test
5	79	17	47	143	0.000***
10	80	21	42	143	0.000***
15	80	28	35	143	0.000***
20	81	30	32	143	0.000***

*** = significant with an error probability of 1%, ** = significant with an error probability of 5%,
* = significant with an error probability of 10%.

The effect sizes of the learning process can be described with either the Pearson correlation coefficient r (Fritz, Morris & Richler, 2012) or using Cohen's d (Cohen, 1992; Cohen, 1988). Pearson's correlation coefficient examines the strength of the correlation between two samples. Cohen's d, on the other hand, considers the expected values of two distributions – the further apart they are, the higher it is. In this way, the first 5 (10/15/20) forecasting rounds can be compared with the last 5 (10/15/20) forecasting rounds. Whereas Pearson's correlation coefficient r corresponds to strong effects according to its categorization by Cohen (1992), Cohen's d shows the average effect sizes of the learning process (Table 5).

Table 5: Effect sizes of the learning processes according to Pearson's r and Cohen's d

Comparison	Pearson's correlation coefficient r	Cohen's d
First 5 forecasting rounds compared to the last 5 rounds	0.57	0.73
First 10 forecasting rounds compared to the last 10 rounds	0.54	0.65
First 15 forecasting rounds compared to the last 15 rounds	0.50	0.58
First 20 forecasting rounds compared to the last 20 rounds	0.49	0.56

However, it is also shown that their experiences only convince a part of the subjects to give up their aversion to algorithms. Even at the end of the 40 rounds of forecasting, just under half of the subjects still show no desire to use the algorithm. On average in rounds 36-40, just below 49% of decisions made are still against the algorithm. At this point in time, the subjects must have realized that their intuitive share price forecasts are far inferior to those of the algorithm, but they decline to use it nevertheless.

As a phenomenon, algorithm aversion has a certain similarity to overconfidence. Learning effects lead to a more realistic estimation of the subject's abilities and thus to a decrease in algorithm aversion. However, the phenomenon of algorithm aversion obviously contains additional aspects which cannot be rectified by the gradual recognition of the superior performance of an algorithm. Among many subjects, their reservations towards the algorithm remain even when they have learned through their own experience that foregoing the algorithm is not in their financial interests.

4. Summary

We experimentally examine the persistence of aversion towards algorithms in relation to learning processes. When subjects have to decide whether they should let an algorithm do a task for them or whether they would rather do it themselves, a possible overestimation of their own competence can lead to rejection of the algorithm. Overconfidence can, however, be tempered by a learning process. Repeated tasks, constant feedback and financial incentives can contribute towards subjects gradually learning to better estimate their own abilities. We are interested in the question of whether such learning processes can also contribute to a reduction of algorithm aversion.

In the experiment, the subjects are asked to make share price forecasts (the price will rise or the price will fall). In 40 rounds of forecasting they can either trust their own assessment or put their faith in a forecasting computer (algorithm). Intuitive forecasts usually turn out to be less successful than the algorithm. The payments made to the subjects depend on the success of their forecasts – regardless of whether the forecasts are their own or whether they are made by the algorithm. After each round, the forecasting results are presented. The subjects can see how much payment they have received and how much they would have obtained if they had trusted the algorithm from the very beginning.

Many subjects recognize as early as the first ten rounds of forecasts that their intuitive share price forecasts are clearly inferior to those of the algorithm. They exhibit an increasing readiness to trust in the algorithm. Regression analysis and the Wilcoxon signed-rank test both show that a learning process can significantly weaken a tendency towards algorithm aversion. With the aid of Pearson's r and Cohen's d, it can be seen that the learning process exhibits a moderate effect size. However, it is also shown that in a considerable part of the subjects there is no weakening of algorithm aversion even over 40 rounds of forecasts.

References

Anderson, A. (2013). Trading and Under-Diversification, *Review of Finance*, 17(5), 1699–1741.

Anzanello, M. J. & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions, *International Journal of Industrial Ergonomics*, 41(5), 573-583.

Barber, B. M. & Odean, T. (2001). Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment, *Quarterly Journal of Economics*, 116(1), 261-292.

Barber, B. M. & Odean, T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors, *The Journal of Finance*, 55(2), 773-806.

Becker, O., Leitner, J. & Leopold-Wildburger, U. (2009). Expectation formation and regime switches, *Experimental Economics*, 12(3), 350-364.

Beketov, M., Lehmann, K. & Wittke, M. (2018). Robo Advisors: quantitative methods inside the robots, Journal of Asset Management, 19, 363–370.

Bhatia, A., Chandani, A. & Chhateja, J. (2020). Robo advisory and its potential in addressing the behavioral biases of investors — A qualitative study in Indian context, *Journal of Behavioral and Experimental Finance*, 25.

Castelo, N., Bos, M. W. & Lehmann, D. R. (2019). Task-dependent algorithm aversion, *Journal of Marketing Research*, 56(5), 809-825.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed), Hillsdale, N.J., L. Erlbaum Associates.

Cohen, J. (1992). A power primer, *Psychological bulletin*, 112(1), 155-159.

D'Acunto, F., Prabhala, N. & Rossi, A. G. (2019). The Promises and Pitfalls of Robo-Advising, *The Review of Financial Studies*, 32(5), 1983–2020.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them, *Management Science*, 64(3), 1155-1170.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err, *Journal of Experimental Psychology: General*, 144(1), 114-126.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S. & Peijnenburg, K. (2016). Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence, *Journal of Financial Economics*, 119, 559-577.

Efendić, E., Van de Calseyde, P. P. & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions, *Organizational Behavior and Human Decision Processes*, 157, 103-114.

Erlei, A., Nekdem, F., Meub, L., Anand, A. & Gadiraju, U. (2020). Impact of Algorithmic Decision Making on Human Behavior: Evidence from Ultimatum Bargaining, *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 8(1), 43-52.

Filiz, I., Nahmer, T. und Spiwoks, M. (2019). Herd behavior and mood: An experimental study on the forecasting of share prices, *Journal of Behavioral and Experimental Finance*, 24, 1-10.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics*, 10(2), 171–178.

Fritz, C. O., Morris, P. E. & Richler, J. J. (2012). Effect Size Estimates: Current Use, Calculations, and Interpretation, *Journal of Experimental Psychology: General*, 141(1), 2–18.

Frydman, C. & Camerer, C. F. (2016). The Psychology and Neuroscience of Financial Decision Making, *Trends in Cognitive Sciences*, 20(9), 661-675.

Gilovich, T., Vallone, R. & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences, *Cognitive psychology*, 17(3), 295-314.

Goetzmann, W. N. & Kumar, A. (2008). Equity Portfolio Diversification, *Review of Finance*, 12(3), 433-463.

Hibbert, A. M., Lawrence, E. R. & Prakash, A. J. (2012). Can Diversification Be Learned?, *The Journal of Behavioral Finance*, 13(1), 38-50.

Jung, D., Dorner, V., Glaser, F. & Morana, S. (2018). Robo-Advisory - Digitalization and Automation of Financial Advisory, *Business & Information Systems Engineering*, 60(1), 81-86.

Köbis, N. & Mossink, L. D. (2020). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry, *Computers in Human Behavior*, 114, 1-13.

Ku, C. Y. (2020). When AIs Say Yes and I Say No: On the Tension between AI's Decision and Human's Decision from the Epistemological Perspectives, *Információs Társadalom*, 19(4), 61-76.

Kudryavtsev, A., Cohen, G. & Hon-Snir, S. (2013). "Rational" or "Intuitive": Are Behavioral Biases Correlated Across Stock Market Investors?, *Contemporary Economics*, 7(2), 31-53.

Meub, L., Proeger, T., Bizer, K. & Spiwoks, M. (2015). Strategic coordination in forecasting - An experimental study, *Finance Research Letters*, 13(1), 155-162.

Prahl, A. & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted?, *Journal of Forecasting*, 36(6), 691-702.

Proeger, T. & Meub, L. (2014). Overconfidence as a Social Bias: Experimental Evidence, in: *Economics Letters*, 122(2), 203-207.

Roberts, H. V. (1959). Stock market ''patterns'' and financial analysis: Methodological suggestions, *Journal of Finance*, 1(14), 1–10.

Rossi, A. G. & Utkus, S. P. (2020). Who Benefits from Robo-advising? Evidence from Machine Learning, SSRN Working Paper, https://ssrn.com/abstract=3552671 or http://dx.doi.org/10.2139/ssrn.3552671

Rühr, A., Streich, D., Berger, B. & Hess, T. (2019). A Classification of Decision Automation and Delegation in Digital Investment Systems, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, S. 1435-1444.

Singh, I. & Kaur, N. (2017). Wealth Management Through Robo Advisory, *International Journal of Research - Granthaalayah*, 5(6), 33-43.

Uhl, M. W. & Rohner, P. (2018). Robo-advisors versus traditional investment advisors: An unequal game, *The Journal of Wealth Management*, 21(1), 44-50.

Wärneryd, K.-E. (2001). Stock-market psychology, Cheltenham: Edward Elgar.

Wright , T. P. (1936). Factors affecting the cost of airplanes, *Journal of the Aeronautical Sciences*, 3(4), 122-128.

Zielonka, P. (2004). Technical analysis as the representation of typical cognitive biases, *International Review of Financial Analysis*, 13, 217–225.

Appendix 1: Instructions for the game

The Game

In this game you are requested to make forecasts on the future trend of a share price. You will forecast the price movements of a share (share Z) in 40 periods. However, you do not predict the exact price of the stock, you only forecast whether it will rise or fall. The price of share Z is always moving, it never remains unchanged. It rises or it falls.

The price of share Z in € at the point in time t (**K_t**) is always determined by four fundamental influencing factors (**A_t**, **B_t**, **C_t** and **D_t**) and a random influence (**ε_t**). The fundamental influencing factors are announced before every round of forecasting. The subjects are also aware of the specific influence the fundamental data has on the share price.

The price (**K_t**) of share Z is formed as follows:

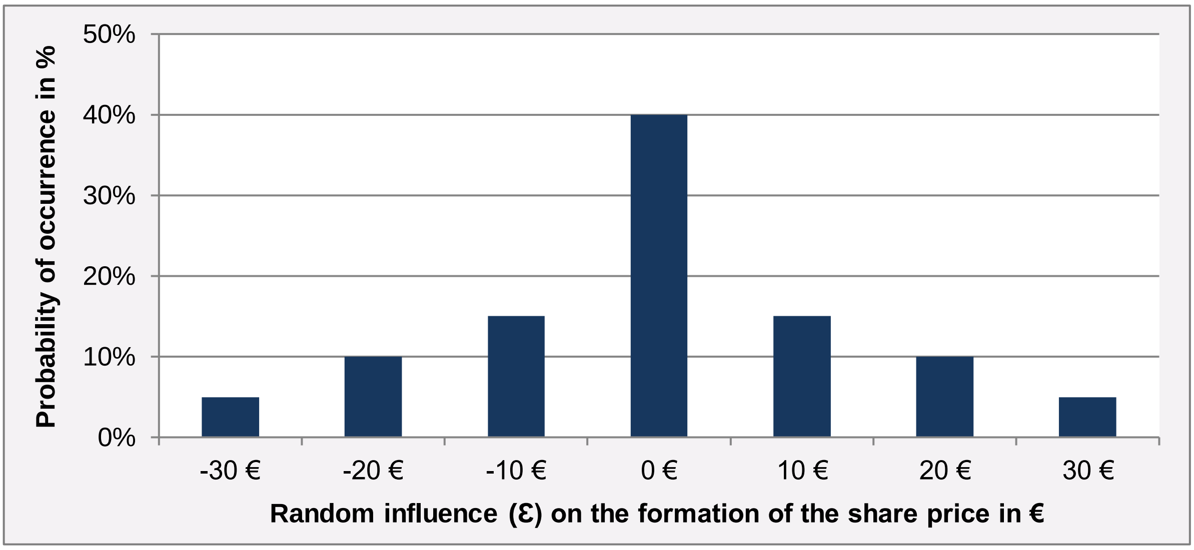
$$K_t = 32 \cdot A_t + 1 \cdot B_t - 18 \cdot C_t + 44 \cdot D_t + \varepsilon_t$$

The fundamental influencing factors (A_t, B_t, C_t and D_t) would, without the random influence ε_t, lead to a price change of between –€10 and €0 or between €0 and +€10 in every period.

If the fundamental influencing factors generally develop favorably, without the random influence ε_t there would always be a rise of the share price between €0 und €10. This means: *€0 < (ΔK_t - ε_t) < +€10.* If, however, the fundamental influencing factors generally develop unfavorably, without the random influence ε_t there would always be a fall in the share price of between €0 und €10. This means: *€0 > (ΔK_t - ε_t) > –€10.*

The random influence ε_t has an expected value of 0 and is distributed as follows: with a probability of 40%, the random influence ε_t is equal to zero ($\varepsilon_t = 0$). With a probability of 15% each, the random influence ε_t obtains a value of –€10 or +€10. With a probability of 10% each, the random influence ε_t obtains a value of –€20 or +€20. With a probability of 5% each, the random influence ε_t obtains a value of –€30 or +€30 (Figure 1).

Figure 1: Distribution probability of the random influence ε_t



In each round of forecasting you have the opportunity to make your own assessment (the price rises or the price falls), or to delegate the decision to a forecasting computer. In 70% of cases, the forecasting computer estimates the future price of stock Z correctly.

Procedure

After reading the instructions and answering the test questions, you see the history of stock Z over the last 40 periods as well as a detailed chart of the price of Z during the last 10 periods. In addition, you will receive the figures of the fundamental data for the next period. You will be asked to forecast the trend of the share price in the next period. After making your forecast you will see what actually happens to the price of the share Z in the next period and receive the results of your prediction. A total of 40 rounds are played. Before every round you see the course of Z from period 1 to the current period as well as a detailed chart of the price of Z during the last 10 periods. In addition, you will receive the figures of the fundamental data for the next period.

Payment

For every successful share price forecast you receive €0.50. A forecast is considered successful and is rewarded accordingly when it correctly predicts the actual direction of the share price. In total you can earn up to €20. Payment is made at the end of the experiment.

Information

- Please remain quiet during the experiment
- Please do not look at your neighbor’s screen
- Apart from a pen and a pocket calculator, **no** aids are permitted (smartphones, smart watches etc.)

Appendix 2: Test questions

Test question 1: Which alternatives do you have when making your forecasts?

a) I can only use my own forecast.
b) I can either follow the algorithm or make my own forecast. *(correct)*
c) I can either follow the algorithm, make my own forecast, or ask other people in the room.

Test question 2: What is the success rate of the algorithm?

a) 40%
b) 50%
c) 70% *(correct)*

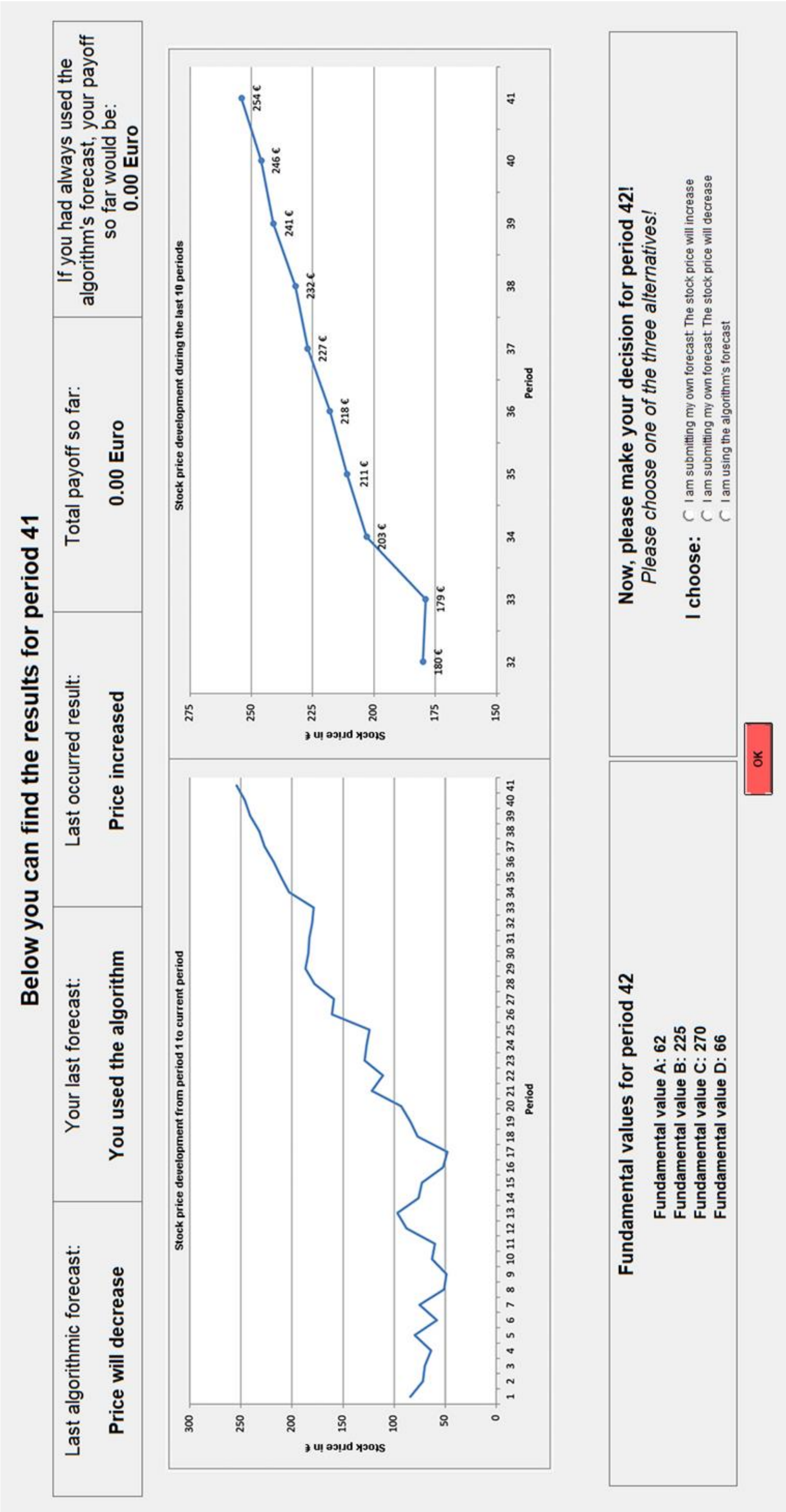
Test question 3: How much is the payment for a successful forecast?

a) €0.00
b) €0.50 *(correct)*
c) €1.00

Test question 4: How much is the payment for an unsuccessful forecast?

a) €0.50
b) €1.00
c) €0.00 *(correct)*

Appendix 3: Screen



Appendix 4: Variations in the price movements

Forecasting round	Variant A	Variant B	Variant C	Variant D
1	267	264	254	255
2	275	273	273	284
3	284	282	277	292
4	272	280	265	289
5	299	278	252	307
6	296	316	261	302
7	294	315	256	310
8	313	333	289	299
9	311	330	287	328
10	349	356	283	326
11	357	364	252	335
12	376	381	280	304
13	364	388	288	283
14	353	377	307	321
15	351	376	306	316
16	389	410	304	305
17	398	408	281	302
18	396	406	278	300
19	405	395	285	308
20	413	414	314	337
21	381	432	323	346
22	410	421	322	335
23	428	408	330	344
24	436	417	308	332
25	425	426	336	310
26	423	434	335	308
27	422	423	334	327
28	421	421	332	325
29	430	418	339	333
30	438	427	358	341
31	446	444	323	349
32	454	431	331	325
33	463	398	310	304
34	461	392	313	312
35	450	400	322	311
36	448	408	311	320
37	427	384	308	317
38	408	392	326	314
39	387	370	324	343
40	366	355	333	349

In order to prevent distortion of the results due to conferring between subjects who take part in the experiment at different points in time, four different variants of price movements were used in the experiment.