ALGORITHM AVERSION AND OTHER CAUSES OF BIAS IN DECISION BEHAVIOR

Studies on Algorithm Aversion, Capital Market Forecasting, and Price Dispersion

DISSERTATION
in order to acquire the doctoral degree
from the Faculty of Economic Sciences
at University of Göttingen

submitted by

MARCO LORENZ

Göttingen, 2022

First supervisor:	PROF. DR. KILIAN BIZER
Second supervisor:	PROF. DR. MARKUS SPIWOKS
Third supervisor:	PROF. DR. HOLGER RAU

Disclaimer

Der Verfasser dieser Dissertation hat während seines Promotionsstudiums als Doktorand bei der Volkswagen Aktiengesellschaft gearbeitet. Ergebnisse, Meinungen und Schlüsse dieser Dissertation sind nicht notwendigerweise die der Volkswagen Aktiengesellschaft.

The author of this dissertation worked as a doctoral student at Volkswagen Aktiengesellschaft during his studies. The results, opinions and conclusions expressed in this thesis are not necessarily those of Volkswagen Aktiengesellschaft.

Table of Contents

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Chapter VI

Chapter VII

Chapter VIII

Chapter IX

Appendix

Chapter I

Introduction

"[Political economy] does not treat the whole of man's nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth, and who is capable of judging the comparative efficacy of means for obtaining that end. (…) Just in the same manner does Political Economy presuppose an arbitrary definition of man, as a being who inevitably does that by which he may obtain the greatest amount of necessaries, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained." (Mill, 1836)

The neoclassical economic theory remains very influential to this day (Daxhammer & Facsar, 2017). It builds on approaches from classical economics (e.g., Smith, 1776) and attempts to explain the economic reality with the help of models. One of these models is the *Homo Oeconomicus* model. It states that economic subjects always pursue the goal of maximizing their pecuniary benefits in all of their actions. Likewise, it is also assumed that they process all relevant information at all times and proceed in a strictly rational manner (Söllner, 2021).

Based on the *Homo Oeconomicus* model, many prominent concepts were developed, such as the neoclassical market model. It states that markets always tend towards equilibrium over the long term. If all consumers in a market behave like *Homo Oeconomicus*, and a seller asks a price higher than the equilibrium price, a fully informed consumer would no longer buy from this seller, which would force the seller out of the market (Samuelson, 1941).

In their entirety, these models are of great importance for research and teaching in economics. Many theories that up until today continue to influence political decisions regarding, for example, fiscal policy or the labor market are based on the ideas of the neoclassical line of thought (e.g., Corden & Findlay, 1975; Friedman, 1948; Schumpeter, 1942; Hicks, 1939).

However, the assumptions of neoclassicism are increasingly being called into question, since they are often in sharp discrepancy with observations from practice. This is mainly due to the fact that the model is based on a number of prerequisites which are not fulfilled in practice, not least due to physical limitations of the actors involved. For a functioning equilibrium price, for example, information about all available choices and their consequences would have to be freely available to all actors at all times (Walker, 1993), who then optimally process them in fractions of a second (Ötsch, 2019), and immediately translate them into actions without adjustment processes (Bridel, 1997).

Moreover, it is argued that important economic events that are not directly reflected in the market are not adequately accounted for in neoclassical models. These include, but are not limited to, unpaid labor (Ötsch, 2019, p. 206), firm characteristics (Ötsch, 2019, p. 309), negotiations and other interactions outside the market (Ötsch, 2019, p. 311), or technological progress (Ötsch, 2019, p. 324).

Neoclassicism also seems to disregard many important factors influencing economic reality at the level of individuals. For example, it is criticized that social aspects and demographic characteristics do not find a place in the neoclassical model, although they play an important role in practice (Wolfson, 1994). Furthermore, it is criticized that the preferences of the actors are described as mutually independent and unchanged in the long run as well as free of goal conflicts, spontaneity, and social influences (Albert, 2009; Fullbrook, 2004).

In neoclassical theory, the price is the sole determinant of who we do business with. In practice, however, trust in the business partner also seems to play a crucial role (Haas & Deseran, 1981). In the model, the parties involved do nothing but react to price changes. Strategies, risks, and proactive decisions do not exist in the neoclassical model (Kay, 1984). Likewise, it is assumed that people always act transparently and honestly far from strategic behavior, which is also not to be observed in practice (Gerschlager, 2001).

Last but not least, it is criticized that the model is nowadays misappropriated by laymen and, for example, also used in the case of imperfect competition (Ötsch, 2019). Although market structures in the neoclassical sense are de facto man-made and actively implemented by institutions, they are often described as the natural and inherently human response to almost all economic problems (Williams, 1999). In particular, scholars criticize that the neoclassical school of thought is applied to macroeconomic phenomena such as economic crises or voluntary unemployment which cannot be explained adequately using the model without modifications (Lo, 2005; Franz, 2004).

This has led to the emergence of numerous amendments that attempt to reconcile neoclassical theory with recent empirical evidence (e.g., Burdett & Judd, 1983). Other scholars argue that entirely new theories are needed in order to understand economic reality, such as behavioral economics. Behavioral economics is about interpreting economic events against the background of psychological peculiarities, such as the influence of cognitive barriers or social behavior on the decisions of the actors (Mullainathan & Thaler, 2000). Thus, this dissertation contributes to the controversial debate whether the neoclassical model is still tenable with adjustments or should be replaced by more sophisticated approaches. To this end, three specific fields of research are used to examine the extent to which the neoclassical model is appropriate for their further exploration.

A young field of research that shows deviations from the self-interest, full information, and rationality assumptions of the neoclassical *Homo Oeconomicus* model is algorithm aversion. This research area aims to expand our understanding of how humans interact with novel innovations that solve complex problems through structured mathematical processes, so-called algorithms. Önkal et al. (2009) found in an experimental study that the same recommendations that can help subjects maximize their utility are followed to a lesser extent when subjects are led to believe that they are coming from an algorithm and not from a knowledgeable human. Even though the concept of algorithm aversion was not yet established at the time, this finding was subsequently often interpreted as evidence of algorithm aversion (De Baets & Harvey, 2020; Longoni, Bonezzi & Morewedge, 2019; Prahl & van Swol, 2017).

Six years later, Dietvorst, Simmons & Massey (2015) found that errors made by an algorithm lead to a greater loss of trust than errors made by a human being (as has also been shown in, for example, Renier, Schmid Mast & Bekbergenova, 2021; Berger et al., 2021; Prahl & Van Swol, 2017). In an economic experiment, participants could delegate a forecasting task to either an algorithm or a human. Subjects who were able to observe the algorithm making mistakes in preliminary rounds were much more reluctant to tie their incentives to the algorithm, even when they saw that the algorithm clearly outperformed the human. To describe this phenomenon, the term "algorithm aversion" was coined in its original form.

Further studies have shown that even before observing an algorithm perform, many subjects prefer inferior human judgement over a superior algorithm. Even when there is clear evidence that without the use of an algorithm, subjects achieve poorer results in the long run, have to bear higher costs, or enjoy less convenience, many of them consciously forego the use of algorithms (Mahmud et al., 2022; Kawaguchi, 2021; Burton, Stein & Jensen, 2020).

A decision against an algorithm that is known to outperform all alternatives can be in contradiction to the assumptions of full information and rationality of the *Homo Oeconomicus* model. It reduces the expected utility of the decision maker and can consequently be viewed as a behavioral anomaly. To broaden our understanding of this interesting finding, the first five studies in this dissertation (Chapters 2 to 6) deal with the phenomenon of algorithm aversion along with its causes and consequences.

Algorithm aversion is by no means the only research field in which the behavior of the actors contradicts the *Homo Oeconomicus* model. The field of forecasting also makes an interesting contribution to the debate about the usefulness of the neoclassical model in this regard. More than eighty years ago, Ogburn (1934) established that human forecasters tend to systematically underestimate the variability of reality. Even forecasts drawn up by experts are regularly based too closely on the status quo and their accuracy rarely exceeds that of a naïve forecast, that is, predicting that everything will remain as it is (Kunze et al., 2017; Beechey & Österholm, 2014; Tabak & Feitosa, 2008). Furthermore, it has been shown that forecasts are often biased, i.e., have systematic errors (Baghestani & Marchon, 2012; Fraser & McDonald, 1993; Lakonishok, 1980). Nevertheless, important economic policy decisions are made on the basis of these inadequately created forecasts.

A fully informed *Homo Oeconomicus* should be able to notice the discrepancy between the own expectations and reality and should correct the behavior in the subsequent forecasts. However, in reality, even professional forecasters tend to make flawed prognoses over long periods of time. In a further part of this dissertation, biases in human forecasting behavior are therefore examined using the example of interest rate and stock market forecasts in various different markets across the globe (Chapter 7 and 8).

The third and final research area covered in this dissertation is the empirical study of price dispersion (Chapter 9). Empirical evidence suggests that the assumption that markets tend towards equilibrium is also broken in practice. In real markets, large persistent price dispersions can be observed frequently even for entirely homogenous goods (Baye, Morgan & Scholten,

2004; Brown & Goolsbee, 2002; Pratt, Wise & Zeckhauser, 1979). The eighth study of this dissertation therefore contributes to the debate on the role of the neoclassical market model.

Taken as a whole, the results of this dissertation point in a clear direction. The behavior of subjects in economic decision making, forecasting, and markets, is usually at odds with what the models of neoclassical school of thought suggest. Neither a *Homo Oeconomicus* nor a neoclassical market can be observed in practice. Today's widespread use of the neoclassical model for important economic policy decisions should therefore be reconsidered, as it may lead to suboptimal conclusions and decisions. Modern approaches, such as behavioral economics, often seem to be much better suited to explaining economic reality.

First contribution: Reducing Algorithm Aversion through Experience[1]

Previous studies suggest that algorithm aversion could be partly due to decision-makers overestimating their own forecasting abilities (overconfidence bias). By overestimating their own abilities, they underestimate the potential of delegating a task to a specialized algorithm rather than performing it themselves. Therein lies an opportunity. In other contexts, it has been shown that overconfidence can be reduced as soon as subjects can recognize that their prognoses fall short of their own expectations. With this experimental study, it is therefore examined whether algorithm aversion can be reduced through learning effects that lead to a better understanding of one's own forecasting competence.

The subjects are asked to make a forecast whether a share price will rise or fall in each of 40 rounds of the game. They receive a lot of information related to the task. They get an insight into the share price formation mechanism and the share price history, as well as information about four fundamental influencing factors on the share price. In each round, the subjects can decide whether they want to make their own forecasts or delegate the decision to an algorithm that has a success rate of 70%. Regardless of this decision, every correct prediction is rewarded, whether it comes from the algorithm or from the participants themselves. The learning effects are made possible by leaving the task unchanged for 40 game rounds and subjects receiving feedback on their own performance as well as on the performance of the algorithm after each round.

Although subjects cannot exceed the algorithm's probability of success even with optimal use of all available information, they delegate the forecasting task to the algorithm in only 45.87% of the game rounds. They show a strong rejection attitude towards the algorithm. In this study, human behavior is clearly not in line with the *Homo Oeconomicus* model.

However, as the length of the game increases, the subjects rely more frequently on the algorithm. While in the first five rounds of forecasting, only 25.17% of the decisions are in favor of using the algorithm, in the last five rounds it is 51.05%. The proportion of decisions in favor of the algorithm

[1] The following summaries are intended to provide a brief descriptive presentation of the research questions, research designs, and key findings. For this reason, extensive presentations of the linked literature are omitted here. The detailed references can be found in the full texts of the individual studies in the respective chapters of this thesis.

increases, especially in the course of the first 20 rounds. Learning to use an algorithm seems to show the characteristics of a degressive learning curve with a flattening course.

The results show that learning effects are suitable for reducing algorithm aversion. Particularly at the beginning, individuals seem ready to be convinced of the advantages of the algorithm. After a few interactions, however, most of them have developed clear preferences and hardly move away from them. The potential to reduce algorithm aversion through learning effects is therefore limited.

Second contribution: The Tragedy of Algorithm Aversion

The overarching goal of the experimental study "The Tragedy of Algorithm Aversion" is to examine the extent of algorithm aversion in decision-making situations with consequences of varying severity. In addition, an overview of definitions of the term algorithm aversion is developed and seven essential differences in understanding the term are revealed. It is argued that from the perspective of behavioral economics, one can only speak of algorithm aversion in the sense of a behavioral anomaly when an algorithm is the superior option for carrying out an activity, the decision-makers are aware of this fact, and they nevertheless decide against using the algorithm.

To answer the research question, an economic experiment with a between-subjects design is carried out. The subjects are each presented with one of six decision-making situations (e.g., the selection of cooking recipes for the users of an app). They first state how serious they consider the consequences of this decision-making situation for the end customer. Then they decide whether to delegate the task to an algorithm or to a human specialist. In each of the six scenarios, the probabilities of success are identical. It is announced that the algorithm has a 70% success rate, while the human expert has a 60% success rate. The prerequisites are therefore matched to the recommended definition of algorithm aversion. The subjects are informed that they will only receive a bonus payment if the task is successfully completed.

The results are astonishingly clear. Overall, 39.16% of the subjects decide against using the algorithm despite its superiority and despite the higher expected value of their remuneration, meaning that they don't act like *Homo Oeconomicus* at all. On the contrary: due to the heuristics of algorithm aversion, in almost two-fifths of all cases the easily recognizable rational strategy is not chosen. It turns out that emphasizing the statistical superiority of an algorithm is not enough to dispel reservations.

The framing of the decision-making situations also has an impact on the extent of the algorithm aversion. While 29.17% of the subjects decide against the algorithm in situations with less serious consequences, it is even 49.30% in situations with serious consequences. Consequently, the aversion to algorithms that leads to deviating from the rational strategy is particularly pronounced in situations with serious consequences. This is remarkable because it is precisely in these situations that failure to follow a rational strategy can cause the greatest damage.

Third contribution: Comparing Different Kinds of Influence on an Algorithm in Its Forecasting Process and Their Impact on Algorithm Aversion

Failure to use superior algorithms can be strongly related to a human desire for control. Previous research has shown that willingness to use an algorithm can increase significantly when decision makers know they can later adjust the forecasts produced by an algorithm upward or downward. However, this always harbors the risk that the overall forecast quality will decline as a result of the subsequent adjustments. In this study, it is therefore examined whether the reduction in algorithm aversion can also be achieved by giving subjects another opportunity to influence the prognosis process, that is before making the prognoses.

To answer the research question, three treatments are designed in the context of a share price forecast. In the first treatment, the subjects have no way of influencing anything at all (control group). In the second treatment, as in previous studies, subjects are allowed to adjust the predictions of the algorithm by a few points afterwards. In the third treatment, they have the opportunity to change an input variable that the algorithm uses to create its forecasts. In all three treatments, subjects have to decide once at the beginning whether they would like to carry out the forecasting task by themselves or whether the algorithm should carry out the task for them throughout the entire experiment. Performance-based remuneration serves as an incentive for the subjects.

The results confirm that the possibility of influencing the algorithmic output is suitable for significantly reducing algorithm aversion. Without the possibility of influencing, only 44.23% of the subjects decide to use the algorithm. With the option to adjust the forecasts afterwards, it is already 69.23%. The ability to influence the input of the algorithm results in 58.49% of the participants opting for the algorithm. The approach is therefore not suitable for reducing algorithm aversion to the same extent as the possibility of influencing the output.

The study expands our understanding of the effects of influencing an algorithm. Giving decision-makers the final say in dealing with an algorithm proves to be a particularly effective way of reducing algorithm aversion. Influencing the algorithm at an earlier stage is also effective, but does not increase the willingness to use it to the same extent.

Fourth contribution: Impact of the Decoy Effect on Algorithm Aversion

Algorithm aversion is a major barrier to establishing powerful innovation based on automated decision-making. Robo-advisors are a case in point. A robo-advisor manages a client's assets by trying to invest them in a way that generates high returns with low risk. Numerous studies have shown that robo-advisors can outperform humans in this endeavor. Nevertheless, many users find it difficult to entrust their assets to an algorithm instead of a human being due to algorithm aversion. Thus, finding ways to reduce algorithm aversion is an important task. This study examines whether the decoy effect is suitable for reducing algorithm aversion and thus facilitating the introduction of innovations.

The decoy effect exploits a heuristic in human behavior. A decision between two alternatives that are superior to each other in different dimensions is influenced by adding a third alternative. The trick is that this third option (the so-called decoy) is very similar to and clearly inferior to one option (the target), but not similar to the other option (the competitor). It has been shown that adding a decoy to a choice set increases the probability that the target will be chosen instead of the competitor.

In this experimental study, an investment game with two treatments is played in the context of robo-advisors. The first treatment serves as a control group. Here, subjects have to decide whether to entrust their assets to Robo-Advisor A (target) or to manage them themselves (competitor). In the second treatment, they can also choose Robo-Advisor B (decoy) in addition to the independent asset management and Robo-Advisor A. The decoy is designed exactly like the target. The only exception is that it has a lower success rate in order to create a decoy effect. The game lasts ten rounds. In each new round, subjects make a new decision between the two (control group) or three (decoy treatment) possible choices.

First, it turns out that the investment decision in the focus of this study is also massively affected by algorithm aversion. In 56.19% of the cases, subjects decide against the robo-advisors, although their use would lead to a significantly higher remuneration. However, there is hardly any difference in the extent of algorithm aversion between the two treatments. When a second, less powerful robo-advisor is offered as a decoy to influence decision behavior, this only increases the proportion of the target robo-advisor from 42.25% to 42.63%. Likewise, the behavior after observing errors of an algorithm or a subject itself is not influenced significantly by the presence of a decoy. The decoy effect is not conducive to reducing Algorithm Aversion.

Fifth contribution: Algorithm Aversion as an Obstacle in the Establishment of Robo Advisors

In the previous study (Chapter 5), it is shown that algorithm aversion can be a major obstacle for robo-advisor providers. Subjects decide against an algorithm, even though it has advantages in completing a task and, consequently, a higher probability of achieving good results. Subjects knowingly choose to face the consequences of their risky decision not to use the robo-advisor. This study raises the question of whether the results are different if it is not the users themselves but third parties who have to bear the consequences of a decision. Ultimately, this carries the risk that the person directly affected by the decision will subsequently demand justification. This problem could, in theory, easily be circumvented by shifting the responsibility to an algorithm.

In this study, subjects make decisions between different stock portfolios. Their goal is to increase the risk-adjusted return of their portfolios. They can put together stock portfolios themselves or have them put together by a robo-advisor. There are two treatments. In the first treatment, the remuneration achieved is paid directly to the subject who selected the respective portfolio. In the second treatment, the remuneration is paid to another participant. Here, the subjects make decisions on behalf of others.

Once again, algorithm aversion can be observed in such a way that in the majority of cases (59.69%) the subjects choose to compile the portfolios themselves, although this reduces the

average remuneration significantly. It can be clearly seen that the subjects in the second treatment are more successful in making good decisions. If their decisions affect the remuneration of third parties, the subjects who do not delegate the task to the robo-advisor are significantly more likely to select the optimal portfolio (41.45% instead of 30.16%).

Surprisingly, the extent of algorithm aversion remains almost unaffected. When the decision is made on behalf of others, the robo-advisor is used 39.69% of the time (instead of 40.94%) to maximize the expected value of the reward. The subjects would not have had to bother or take any risks if they had delegated the decision to the algorithm, which is specialized in compiling stock portfolios. However, algorithm aversion appears to be so persistent that they decide against this path. The results suggest that algorithm aversion is just as persistent when we make decisions for those around us.

The first five studies in this dissertation underscore how persistent algorithm aversion is and the wide variety of tasks it affects. This dissertation has uncovered that algorithm aversion comes into play especially when a decision has serious consequences. It also shows that algorithm aversion is extremely difficult to overcome. Of the approaches investigated, only learning effects (Chapter 2) and subsequent influence on predictions (Chapter 4) have a statistically significant effect. To expand our understanding of novel and increasingly relevant decisions between humans and algorithms, *Homo Oeconomicus* does not seem to be a helpful model.

Sixth contribution: Interest Rate Forecasts in Latin America

The business model of many financial institutions is based on maturity transformation. The term refers to the balancing of diverging intended maturities of debtors and creditors while making a profit from charging different interest rates. If maturity transformation is carried out based on false expectations about the development of future interest rates, this can threaten the solvency of a financial institution. Reliable interest rate forecasts are thus essential for this business model to work.

The accuracy of professional interest rate forecasts by market analysts has already been studied in detail, especially for forecasts regarding for the United States, but also for Asia and Europe. Interest rate forecasts in Latin America, on the other hand, were rarely the focus.

This study closes the research gap by examining 28,451 forecasts for the money markets in Argentina, Brazil, Chile, Mexico, and Venezuela. The forecasts were drawn up by professional analysts and were published between 2001 and 2019 in the journal Latin American Consensus Forecasts. The forecasts were presented monthly and have a horizon of 4 and 13 months. In order to get a clear picture of the forecast quality on the Latin American money markets, a variety of established statistical methods is applied to the data, i.e., the Diebold-Mariano test, the sign accuracy test, the TOTA coefficient, and the unbiasedness test.

A majority of the forecasts perform well in comparison to the naïve forecast and also in the sign accuracy test. This suggests that the forecasts do indeed capture valuable information and that taking the forecasts into account could lead to successes in maturity transformation. On the other

hand, however, most forecasts are biased and subject to the phenomenon of topically oriented trend adjustment. They are therefore geared too closely to current trends.

Another insight is that the forecast quality varies strongly between markets. While in line with most previous studies on other regions, forecasts for Argentina and Venezuela show poor quality, the forecasts for Brazil, Chile and Mexico are significantly better than those for most other countries considered by researchers so far.

Seventh contribution: Sticky Stock Market Analysts

The accuracy of capital market forecasts has long been the subject of scientific research. If it is possible to make accurate forecasts about future developments, an excess return can be achieved on the market based on these forecasts. Actors could then maximize their income by investing in the stocks that will rise in value and selling the stocks that will fall in value. However, previous studies have shown that the quality of capital market forecasts is usually not sufficient to enable a promising active investment strategy.

In this study, the accuracy of forecasts for the German Stock Market Index (DAX), the Dow Jones Industrial Index (DJI), and the Euro Stoxx 50 (SX5E) are analyzed. The forecasts were drawn up by professional analysts and were published in the German business newspaper “Handelsblatt” and the quality broadsheet “Frankfurter Allgemeine Zeitung” between 1992 and 2020. They were made at the end of the previous year and have a horizon of six months and one year.

All three stock market indices display a rising trend in the long run. However, there are also phases in all three indices in which the course falls. It is well known that forecasters like to present conservative prognoses that are based on the long-term trend and only include small changes. It is examined whether falling prices are predicted with sufficient frequency and whether the magnitude of the predicted changes (in terms of standard deviation) approximates the magnitude of the actual changes. In addition, in the form of the prediction-realization diagram, unbiasedness test, and Diebold-Mariano test, established statistical methods are applied.

It turns out that the forecast quality for all three indices is low. The forecasts are highly biased and lag behind actual events in the indices. Moreover, the forecasts for the Euro Stoxx 50 are even significantly worse than the corresponding naïve forecasts. This means that by assuming that everything stays the way it is, you would obtain better results than by following the forecasts. For all three indices, the forecast behavior shows flaws that were scientifically described as early as the 1930s, namely that the variability of reality is systematically underestimated.

Overall, the results from Chapters 7 and 8 suggest that the forecasters’ behavior does not fit the *Homo Oeconomicus* model. Admittedly, not even a *Homo Oeconomicus* can look into the future and predict price developments with certainty. Nonetheless, if trends are systematically overestimated or underestimated over decades or forecasts are too strongly anchored to the status quo, a *Homo Oeconomicus* should recognize this and adjust the forecasting behavior. However, this seems to succeed only in very few exceptional cases.

Eighth contribution: Unicorn, Yeti, Nessie, and Neoclassical Market – Legends and Empirical Evidence

The final contribution of this dissertation (Chapter 9) focuses on another key element of the neoclassical school of thought: its market model. This study examines the extent to which the theory of the neoclassical market model is valid in practice.

At the heart of the neoclassical market model are three concepts: the aggregate supply function, the aggregate demand function, and the equilibrium price at their intersection. While the former two are not observable in practice, empirical research can provide a good picture of the third concept, the equilibrium price. To this end, we evaluate price comparisons that eleven teams of students in Lower Saxony conducted as part of lectures and theses. The students noted prices that different retailers in stationary trade and online trade set for one and the same homogeneous product. The prices were always recorded at the same time and in the same economic area to enable meaningful analysis. The resulting database includes 146 price comparisons for homogeneous products composed of a total of 2,217 prices.

It turns out that no equilibrium price can be observed for 143 out of 146 price comparisons. For many products, there are retailers who charge more than double than the cheapest offer. The percentage price range is 125.96% on average. The average coefficient of variation is 0.20. Both values are far from 0, the value suggested by the theory.

According to neoclassical theory, sellers who charge more than the equilibrium price should not find buyers and should thus be forced out of the market in the long run. This obviously does not happen in practice. There seem to be significantly more influencing factors than just the price that are decisive for transactions to take place. We argue that the steering function of the neoclassical market model and the dominant role it plays in research and teaching should be critically questioned in light of this evident discrepancy between theory and practice.

References

Albert, H. (1972). Ökonomische Theorie als politische Ideologie: Das ökonomische Argument in der ordnungspolitischen Debatte, 3rd Edition, Mohr Siebeck, Tübingen.

Baghestani, H., & Marchon, C. (2012). An evaluation of private forecasts of interest rate targets in Brazil. *Economics Letters*, *115*(3), 352-355.

Baye, M. R., Morgan, J., & Scholten, P. (2004). Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site, *The Journal of Industrial Economics*, *52*(4), 463-496.

Beechey, M., & Österholm, P. (2014). Policy interest-rate expectations in Sweden: a forecast evaluation, *Applied Economics Letters, 21*(13/15), 984-991.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve - Algorithm Aversion and Demonstrating the Ability to Learn, *Business & Information Systems Engineering*, *63*(1), 55-68.

Bridel, P. (1997). Money and General Equilibrium Theory: From Walras to Pareto (1870-1923), Edward Elgar Publishing, Cheltenham.

Brown, J. R., & Goolsbee, A. (2002). Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry, *Journal of Political Economy*, *110*(3), 481-507.

Burdett, K., & Judd, K. L. (1983). Equilibrium Price Dispersion, *Econometrica, 51*(4), 955-969.

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making, *Journal of Behavioral Decision Making*, *33*(2), 220-239.

Corden, W. M., & Findlay, R. (1975). Urban Unemployment, Intersectoral Capital Mobility and Development Policy, *Economica*, *42*(165), 59-78.

Daxhammer, R. J., & Facsar, M. (2017). Behavioral Finance: Verhaltenswissenschaftliche Finanzmarktforschung im Lichte begrenzt rationaler Marktteilnehmer, 2nd Edition, UVK Verlag, Konstanz.

De Baets, S., & Harvey, N. (2020). Using judgment to select and adjust forecasts from statistical models, *European Journal of Operational Research*, *284*(3), 882-895.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err, *Journal of Experimental Psychology, 144*(1), 114-126.

Franz, S. (2004). Grundlagen des ökonomischen Ansatzes: Das Erklärungskonzept des Homo Oeconomicus, *University of Potsdam Working Paper 2004-02,* Potsdam.

Fraser, P. & McDonald, R. (1993). The Efficiency of CAC Stock Price Forecasts: A Survey Based Perspective, *Revue économique*, *44*(5), 991-1000.

Friedman, M. (1948). A Monetary and Fiscal Framework for Economic Stability, *The American Economic Review*, *38*(3), 245-264.

Fullbrook, E. (2004). Descartes' Legacy: Intersubjective Reality, Intrasubjective Theory, In: Davis, J. B., & Marciano, A. (2004): The Elgar Companion To Economics and Philosophy, Edward Elgar Publishing, London.

Gerschlager, C. (2001). Expanding the Economic Concept of Exchange: Deception, Self-Deception and Illusions, Springer Science+Business Media, Dordrecht.

Haas, D. F., & Deseran, F. A. (1981). Trust and Symbolic Exchange, *Social Psychology Quarterly*, *44*(1), 3-13.

Hicks, J. R. (1939). The Foundations of Welfare Economics, *The Economic Journal*, *49*(196), 696-712.

Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business, *Management Science*, *67*(3), 1670-1695.

Kay, N. M. (1984). The Emergent Firm: Knowledge, Ignorance and Surprise in Economic Organisation, Palgrave Macmillan, London.

Kunze, F., Wegener, C., Bizer, K., & Spiwoks, M. (2017). Forecasting European interest rates in times of financial crisis – What insights do we get from international survey forecasts?, *Journal of International Financial Markets, Institutions & Money, 48,* 192-205.

Lakonishok, J. (1980). Stock Market Return Expectations: Some General Properties, *The Journal of Finance*, *35*(4), 921-931.

Lo, A. W. (2005). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis, *Journal of Investment Consulting*, *7*(2), 21-44.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence, *Journal of Consumer Research*, *46*(4), 629-650.

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion, *Technological Forecasting and Social Change*, *175*, 121390.

Mill, J. S. (1836). Essays on Some Unsettled Questions of Political Economy, *London and Westminster Review*, *26*(2), London.

Mullainathan, S., & Thaler, R. H. (2000). Behavioral economics, National Bureau of Economic Research Working Paper No. 7948, Cambridge, https://www.nber.org/papers/w7948 (accessed on August 10th, 2022).

Ogburn, W. F. (1934). Studies in Prediction and the Distortion of Reality, *Social Forces*, *13*, 224-229.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S. & Pollock, A. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments, *Journal of Behavioral Decision Making*, *22*(4), 390-409.

Ötsch, W. O. (2019). Mythos Markt, Mythos Neoklassik: Das Elend des Marktfundamentalismus, Metropolis, Marburg.

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

Pratt, J. W., Wise, D. A., & Zeckhauser, R. (1979). Price Differences in Almost Competitive Markets, *The Quarterly Journal of Economics*, *93*(2), 189-211.

Renier, L. A., Schmid Mast, M., & Bekbergenova, A. (2021). To err is human, not algorithmic – Robust reactions to erring algorithms, *Computers in Human Behavior*, 106879.

Samuelson, P. A. (1941). The Stability of Equilibrium: Comparative Statics and Dynamics, *Econometrica, 9*(2), 97-120.

Schumpeter, J. (1942). Capitalism, Socialism and Democracy, Harper & Brothers, New York City.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations, Strahan & Cadell, London.

Söllner, F. (2021). Die Geschichte des ökonomischen Denkens, 5th Edition, Springer Gabler, Heidelberg.

Tabak, B. M. & Feitosa, M. A. (2008). How Informative are Interest Rate Survey-based Forecasts?, *Brazilian Administrative Review, 5*(4), 304-318.

Walker, D. A. (1993). Walras's Models of the Barter of Stocks of Commodities, *European Economic Review*, *37*(7), 1425-1446.

Williams, D. (1999). Constructing the Economic Space: The World Bank and the Making of Homo Oeconomicus, *Millennium*, *28*(1), 79-99.

Wolfson, M. (1994). Eligo Ergo Sum: Classical Philosophies of the Self in Neoclassical Economics, *History of Political Economy*, *26*(2), 297-325.

Chapter II

Reducing Algorithm Aversion Through Experience

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Published in:
Journal of Behavioral and Experimental Finance, 31, 100524.

Also published in:
Sofia Diskussionsbeiträge, Sonderforschungsgruppe Institutionenanalyse, No. 21-1, Darmstadt, January 2021. (Discussion Paper in German)
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 21-01, Wolfsburg, January 2021. (Working Paper)

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.

Keywords

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

JEL Classification

D83, D84, D91, G17, G41.

Chapter III

The Tragedy of Algorithm Aversion

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Submitted to:
PLOS ONE (first round of revision completed)

Published in:
Sofia Diskussionsbeiträge, Sonderforschungsgruppe Institutionenanalyse, No. 21-2 Darmstadt, January 2021. (Discussion Paper in German)
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 21-02, Wolfsburg, February 2021. (Working Paper)

Abstract

Algorithms already carry out many tasks more reliably than human experts. Nevertheless, some subjects have an aversion towards algorithms. In some decision-making situations an error can have serious consequences, in others not. In the context of a framing experiment, we examine the connection between the consequences of a decision-making situation and the frequency of algorithm aversion. This shows that the more serious the consequences of a decision are, the more frequently algorithm aversion occurs. Particularly in the case of very important decisions, algorithm aversion thus leads to a reduction of the probability of success. This can be described as the tragedy of algorithm aversion.

Keywords

Algorithm aversion, technology adoption, framing, behavioral economics, experiments.

JEL Classification

D81, D91, G41, O33.

1. Introduction

Automated decision-making or decision aids, so-called algorithms, are becoming increasingly significant for many people's working and private lives. The progress of digitalization and the growing significance of artificial intelligence in particular mean that efficient algorithms have now already been available for decades (see, for example, Dawes, Faust & Meehl, 1989). These algorithms already carry out many tasks more reliably than human experts. However, only a few algorithms are completely free of errors. Some areas of application of algorithms have serious consequences in the case of a mistake – such as autonomous driving (cf. Shariff, Bonnefon & Rahwan, 2017), making medical diagnoses (cf. Majumdar & Ward, 2011), or support in criminal proceedings (cf. Simpson, 2016). On the other hand, algorithms are also used for tasks which do not have such severe consequences in the case of an error, such as dating service (cf. Brozovsky & Petříček, 2007), weather forecasts (cf. Sawaitul, Wagh & Chatur, 2012) and the recommendation of recipes (cf. Ueda, Takahata & Nakajima, 2011).

Some subjects have a negative attitude towards algorithms. This is usually referred to as algorithm aversion (for an overview of algorithm aversion see Burton, Stein & Jensen, 2020). Many decision-makers thus tend to delegate tasks to human experts or carry them out themselves. This is also frequently the case when it is clearly recognizable that using algorithms would lead to an increase in the quality of the results.

Previous publications on this topic have defined the term algorithm aversion in quite different ways (Table 1). These different understandings of the term are reflected in the arguments put forward as well as in the design of the experiments carried out. From the perspective of some researchers, it is only possible to speak of algorithm aversion when the algorithm recognizably provides the option with the highest quality result or probability of success (cf. Köbis & Mossink, 2021; Burton, Stein & Jensen, 2020; Ku, 2020; Castelo, Bos & Lehmann, 2019; Dietvorst, Simmons & Massey, 2015). However, other authors consider algorithm aversion to be present as soon as subjects exhibit a fundamental disapproval of an algorithm in spite of its possible superiority (cf. Efendić, Van de Calseyde & Evans, 2020; Niszczota & Kaszás, 2020; Horne et al., 2019; Logg, Minson & Moore, 2019; Rühr et al., 2019; Yeomans et al., 2019; Prahl & Van Swol, 2017).

Another important aspect of how the term algorithm aversion is understood is the question of whether and possibly also how the subjects hear about the superiority of the algorithm. Differing approaches were chosen in previous studies. Dietvorst, Simmons and Massey (2015) focus on the gathering of experience in dealing with an algorithm in order to be able to assess its probability of success in comparison to one's own performance. In a later study, Dietvorst, Simmons and Massey (2018) specify the average error of the algorithm. Alexander, Blinder and Zak (2018) provide exact details on the probability of success of the algorithm, or they refer to the rate at which other subjects used the algorithm in the past.

Table 1: Definitions of algorithm aversion in the literature

Authors	Definition of algorithm aversion
Dietvorst, Simmons & Massey, 2015	"Research shows that evidence-based algorithms more accurately predict the future than do human forecasters. Yet when forecasters are deciding whether to use a human forecaster or a statistical algorithm, they often choose the human forecaster. This phenomenon, which we call *algorithm aversion (…)*"
Prahl & Van Swol, 2017	"The irrational discounting of automation advice has long been known and a source of the spirited "clinical versus actuarial" debate in clinical psychology research (Dawes, 1979; Meehl, 1954). Recently, this effect has been noted in forecasting research (Önkal et al., 2009) and has been called algorithm aversion (Dietvorst, Simmons, & Massey, 2015)."
Dietvorst, Simmons & Massey, 2018	"Although evidence-based algorithms consistently outperform human forecasters, people often fail to use them after learning that they are imperfect, a phenomenon known as *algorithm aversion*."
Castelo, Bos & Lehmann, 2019	"The rise of algorithms means that consumers are increasingly presented with a novel choice: should they rely more on humans or on algorithms? Research suggests that the default option in this choice is to rely on humans, even when doing so results in objectively worse outcomes."
Horne, Nevo, O'Donovan, Cho & Adali, 2019	"For example, Dietvorst et al. (Dietvorst, Simmons, and Massey 2015) studied when humans choose the human forecaster over a statistical algorithm. The authors found that aversion of the automated tool increased as humans saw the algorithm perform, even if that algorithm had been shown to perform significantly better than the human."
Ku, 2019	"(…) "algorithm aversion", a term refers by Dietvorst et al. (Dietvorst et al. 2015) means that humans distrust algorithm even though algorithm consistently outperform humans."
Leyer & Schneider, 2019	"In the particular context of the delegation of decisions to AI-enabled systems, recent findings have revealed a general algorithmic aversion, an irrational discounting of such systems as suitable decision-makers despite objective evidence (Dietvorst, Simmons and Massey, 2018)"
Logg, Minson & Moore, 2019	"(…) human distrust of algorithmic output, sometimes referred to as "algorithm aversion" (Dietvorst, Simmons, & Massey, 2015).[1] "; Footnote 1: "while this influential paper [of Dietvorst et al.] is about the effect that seeing an algorithm err has on people's likelihood of choosing it, it has been cited as being about how often people use algorithms in general."
Önkal, Gönül & De Baets, 2019	"(…) people are averse to using advice from algorithms and are unforgiving toward any errors made by the algorithm (Dietvorst et al., 2015; Prahl & Van Swol, 2017)."
Rühr, Streich, Berger & Hess, 2019	"Users have been shown to display an aversion to algorithmic decision systems [Dietvorst, Simmons, Massey, 2015] as well as to the perceived loss of control associated with excessive delegation of decision authority [Dietvorst, Simmons, Massey, 2018]."
Yeomans, Shah, Mullainathan & Kleinberg, 2019	"(...) people would rather receive recommendations from a human than from a recommender system (...). This echoes decades of research showing that people are averse to relying on algorithms, in which the primary driver of aversion is algorithmic errors (for a review, see Dietvorst, Simmons, & Massey, 2015)."
Berger, Adam, Rühr & Benlian, 2021	"Yet, previous research indicates that people often prefer human support to support by an IT system, even if the latter provides superior performance – a phenomenon called algorithm aversion." (…) "These differences result in two varying understandings of what algorithm aversion is: unwillingness to rely on an algorithm that a user has experienced to err versus general resistance to algorithmic judgment."
Burton, Stein & Jensen, 2020	"(…) algorithm aversion—the reluctance of human forecasters to use superior but imperfect algorithms— (…)"
De-Arteaga, Fogliato & Chouldechova, 2020	"*Algorithm aversion*–the tendency to ignore tool recommendations after seeing that they can be erroneous (…)"
Efendić, Van de Calseyde & Evans, 2020	"Algorithms consistently perform well on various prediction tasks, but people often mistrust their advice. (…) However, repeated observations show that people profoundly mistrust algorithm-generated advice, especially after seeing the algorithm fail (Bigman & Gray, 2018; Diab, Pui, Yankelevich, & Highhouse, 2011; Dietvorst, Simmons, & Massey, 2015; Önkal, Goodwin, Thomson, Gönül, & Pollock, 2009)."

Erlei, Nekdem, Meub, Anand & Gadiraju, 2020	"Recently, the concept of algorithm aversion has raised a lot of interest (see (Burton, Stein, and Jensen 2020) for a review). In their seminal paper, (Dietvorst, Simmons, and Massey 2015) illustrate that human actors learn differently from observing mistakes by an algorithm in comparison to mistakes by humans. In particular, even participants who directly observed an algorithm outperform a human were less likely to use the model after observing its imperfections."
Germann & Merkle, 2020	"The tendency of humans to shy away from using algorithms even when algorithms observably outperform their human counterpart has been referred to as algorithm aversion."
Ireland, 2020	"(…) some researchers find that, when compared to humans, people are averse to algorithms after recording equivalent errors."
Jussupow, Benbasat & Heinzl, 2020	"(…) literature suggests that although algorithms are often superior in performance, users are reluctant to interact with algorithms instead of human agents – a phenomenon known as algorithm aversion"
Niszczota & Kaszás, 2020	"When given the possibility to choose between advice provided by a human or an algorithm, people show a preference for the former and thus exhibit algorithm aversion (Castelo et al., 2019; Dietvorst et al., 2015, 2016; Longoni et al., 2019)."
Wang, Harper & Zhu, 2020	"(…) people tend to trust humans more than algorithms even when the algorithm makes more accurate predictions."
Kawaguchi, 2021	"The phenomenon in which people often obey inferior human decisions, even if they understand that algorithmic decisions outperform them, is widely observed. This is known as algorithm aversion (Dietvorst et al. 2015)."
Köbis & Mossink, 2021	"When people are informed about algorithmic presence, extensive research reveals that people are generally averse towards algorithmic decision makers. This reluctance of "human decision makers to use superior but imperfect algorithms" (Burton, Stein, & Jensen, 2019; p.1) has been referred to as algorithm aversion (Dietvorst, Simmons, & Massey, 2015). In part driven by the belief that human errors are random, while algorithmic errors are systematic (Highhouse, 2008), people have shown resistance towards algorithms in various domains (see for a systematic literature review, Burton et al., 2019)."
Commerford, Dennis, Joe & Wang, 2022	"(…) *algorithm aversion* – the tendency for individuals to discount computer-based advice more heavily than human advice, although the advice is identical otherwise."

In addition, when dealing with algorithms, the way in which people receive feedback is of significance. Can subjects (by using their previous decisions) draw conclusions about the quality and/or success of an algorithm? Dietvorst, Simmons and Massey (2015) merely use feedback in order to facilitate experience in dealing with an algorithm. Prahl and Van Swol (2017) provide feedback after every individual decision, enabling an assessment of the success of the algorithm. Filiz et al. (2021) also follow this approach and use feedback after every single decision in order to examine the decrease in algorithm aversion over time.

Other aspects which emerge from the previous definitions of algorithm aversion in the literature are the reliability of the algorithm (perfect or imperfect), the observation of its reliability (the visible occurrence of errors), access to historical data on how the algorithmic forecast was drawn up; the setting (algorithm vs. expert; algorithm vs. amateur; algorithm vs. subject) as well as extent of the algorithm's intervention (does the algorithm act as an aid to decision-making or does it carry out tasks automatically?).

In our view, the superiority of the algorithm (higher probability of success) and the knowledge of this superiority are the decisive aspects. We only speak of algorithm aversion when subjects are clearly aware that not using the algorithm reduces the expected value of their utility and they do not deploy it nevertheless. A decision against the use of an algorithm which is known to be

superior reduces the expected value of the subject's pecuniary utility and thus has to be viewed as a behavioral anomaly (cf. Frey, 1992; Kahneman & Tversky, 1979; Tversky & Kahneman, 1974).

In decision-making situations which lead to consequences which are not so serious in the case of an error, a behavioral anomaly of this kind does not have particularly significant effects. In the case of a dating service, the worst that can happen is meeting with an unsuitable candidate. In the case of an erroneous weather forecast, unless it is one for seafarers, the worst that can happen is that unsuitable clothing is worn, and if the subject is the recommendation of recipes, the worst-case scenario is a bland meal. However, particularly in the case of decisions which have serious consequences in the case of a mistake, diverging from the rational strategy would be highly risky. For example, a car crash or a wrong medical diagnosis can, in the worst case, result in someone's death. Being convicted in a criminal case can lead to many years of imprisonment. In these serious cases, it is important to be sensible and use an algorithm when it is superior in terms of its probability of success. Can algorithm aversion be overcome in serious situations in order to make a decision which maximizes utility and which, at best, can save a life?

Tversky & Kahneman (1981) show that decisions can be significantly influenced by the context of the decision-making situation. The story chosen to illustrate the problem influences the salience of the information, which can also lead to an irrational neglect of the underlying mathematical facts. This phenomenon is also referred to as the framing effect (for an overview see Cornelissen & Werner, 2014). Irrespective of the actual probability of success, subjects do allow themselves to be influenced. Algorithm aversion can be more or less pronounced in different decision-making contexts. It is possible that subjects who have to decide on the use of an algorithm also take the consequences of their decision into account. This study therefore uses a framing approach to examine whether subjects are prepared to desist from their algorithm aversion in decision-making situations which can have severe consequences. We thus consider whether there are significantly different frequencies of algorithm aversion depending on whether the decision-making situations can have serious consequences or not.

2. Experimental design and hypotheses

In order to answer the research question, we carry out an economic experiment in which the subjects assume the perspective of a businessperson who offers a service to his/her customers. A decision has to be made on whether this service should be carried out by specialized algorithms or by human experts.

In this framing approach, three decision-making situations with potentially serious consequences (Treatment A) and three decision-making situations with significantly less serious effects are compared (Treatment B). In Treatment A it concerns the following services: (1) Driving services with the aid of autonomous vehicles (algorithm) or with the aid of drivers, (2) The evaluation of MRI scans with the help of a specialized computer program (algorithm) or with the aid of doctors, and (3) The evaluation of files on criminal cases with the aid of a specialized computer program (algorithm) or with the help of legal specialists. In Treatment B it concerns the following services: (1) A dating site providing matchmaking with the aid of a specialized computer program (algorithm) or with the support of staff trained in psychology, (2) The selection of recipes for

cooking subscription boxes with the aid of a specialized computer program or the help of staff trained as professional chefs, and (3) The drawing up of weather forecasts with the help of a specialized computer program (algorithm) or using experienced meteorologists (Table 2).

Table 2: Treatments and decision-making situations

Decision-making situation	Treatment
Autonomous driving Evaluation of MRI scans The assessment of criminal case files	A (possibly serious consequences)
Dating service Selection of recipes Drawing up weather forecasts	B (no serious consequences)

The decision-making situations are selected in such a way that the subjects should be familiar with them from public debates or from their own experience. In this way, it is easier for the subjects to immerse themselves in the respective context. Detailed descriptions of the decision-making situations can be viewed in Appendix C.

The study has a between-subjects design. Each subject is only confronted with one of a total of six decision-making situations. All six decision-making situations have the same probability of success: the algorithm carries out the service with a probability of success of 70%. The human expert carries out the service with a probability of success of 60%. The payment structure is identical in both treatments. The participants receive a show-up fee of €2, and an additional payment of €4 is made if the service is carried out successfully. It is only the contextual framework of the six decision-making situations which varies.

First of all, the subjects are asked to assess the gravity of the decision-making situation on a scale from 0 (not serious) to 10 (very serious). This question has the function of a manipulation check - in this way it can be seen whether the subjects actually perceive the implications of the decision-making situations in Treatment A as more serious than those in Treatment B. In the case of autonomous vehicles and the evaluation of MRI scans, it could be a matter of life and death. In the evaluation of documents in the context of criminal cases, it could lead to serious limitations of personal freedom. The three decision-making situations in Treatment A can thus have significant consequences for third parties if they end unfavorably. The situation is different in the case of matchmaking, selecting recipes and drawing up weather forecasts. Even when these tasks cannot be accomplished in a satisfactory way sometimes, the consequences are usually not very severe. A date might turn out to be dull, or one is disappointed by the taste of a lunch, or you are out without a jacket in the rain. None of those things would be pleasant, but the implications in Treatment B are far less serious than those in Treatment A.

A *Homo Oeconomicus* (a person who acts rationally in economic terms) must – regardless of the context – prefer the algorithm to human experts, because it maximizes his or her financial utility. Every decision in favor of the human experts has to be considered algorithm aversion.

Algorithm aversion is a phenomenon which can occur in a wide range of decision-making situations (Burton, Stein & Jensen, 2020). We thus presume that the phenomenon can also be observed in this study. Although the decision-making situations offer no rational grounds for choosing the human experts, some of the participants will do precisely this. Hypothesis 1 is: Not every subject will select the algorithm. Null hypothesis 1 is therefore: Every subject will select the algorithm.

Castelo, Bos & Lehmann (2019) show that framing is suited to influencing algorithm aversion. A dislike for algorithms appears to various degrees in different contexts. Nonetheless, in this study, the algorithm was not recognizably the most reliable alternative, and there is also no performance-related payment for the subjects. In Castelo, Bos & Lehmann (2019), algorithm aversion is therefore not modeled as a behavioral anomaly.

However, we expect that the frame will have an influence on algorithm aversion if the financial advantage of the algorithm is clearly recognizable. Hypothesis 2 is: The proportion of decisions made in favor of the algorithm will vary significantly between the two treatments. Null hypothesis 2 is therefore: The proportion of decisions made in favor of the algorithm will not vary significantly between the two treatments.

In the literature there are numerous indications that framing can significantly influence the decision-making behavior of subjects (cf. Tversky & Kahneman, 1981). If subjects acted rationally and maximized their utility, neither algorithm aversion nor the framing effect would arise. Nonetheless, real human subjects – as the research in behavioral economics frequently shows – by no means act like *Homo Oeconomicus.* Their behavior usually tends to correspond more to the model of bounded rationality put forward by Herbert A. Simon (1959). Human beings suffer from cognitive limitations – they fall back on rules of thumb and heuristics. But they do try to make meaningful decisions – as long as this does not involve too much effort. This kind of 'being sensible' – which is often praised as common sense – suggests that great efforts have to be made when decisions can have particularly severe consequences. The founding of a company is certainly given much more thought than choosing which television program to watch on a rainy Sunday afternoon. And much more care will usually be invested in the selection of a heart surgeon than in the choice of a pizza delivery service.

This everyday common sense, which demands different levels of effort for decision-making situations with different degrees of gravity, could contribute towards the behavioral anomaly of algorithm aversion appearing more seldom in Treatment A (decisions with possible serious consequences) than in Treatment B (decisions with relatively insignificant effects). Hypothesis 3 is thus: The greater the gravity of a decision, the more seldom the behavioral anomaly of algorithm aversion arises. Null hypothesis 3 is therefore: Even when the gravity of a decision-making situation increases, there is no reduction in algorithm aversion.

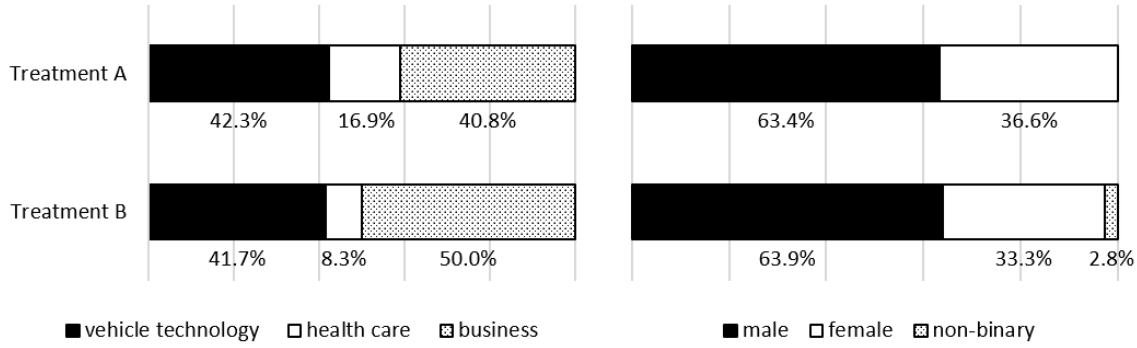
3. Results

This economic 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. A total of 143 students of the Ostfalia University of Applied Sciences take part in the experiment. Of these, 91 subjects are male (63.6%), 50 subjects are female (35%) and 2 subjects (1.4%) describe themselves as non-binary. Of the 143 participants, 65 subjects (45.5%) study at the Faculty of Economics and Business, 60 subjects (42.0%) at the Faculty of Vehicle Technology, and 18 subjects (12.6%) at the Faculty of Health Care. Their average age is 23.5 years.

Of the participants, 71 subjects are in a decision-making situation which has been assigned to Treatment A, while 72 subjects are presented with a decision-making situation which has been assigned to Treatment B. The distribution of the subjects to the two treatments has similarities to their distribution among the faculties as well as their gender. In Treatment A (respectively Treatment B), 42.3% (41.7%) of the subjects belong to the faculty of vehicle technology, while 16.9% (8.3%) belong to the faculty of health care, and 40.8% (50.0%) belong to the faculty of business. In Treatment A (respectively Treatment B), 63.4% (63.9%) of the subjects are male, and 36.6% (33.3%) are female, and 0% (2.8%) are non-binary (Figure 1).

Figure 1: Proportions of the subjects belonging to a faculty and/or gender in Treatments A and B.



The experiment is programmed with z-Tree (cf. Fischbacher, 2007). Only the lottery used to determine the level of success when providing the service is carried out by taking a card from a pack of cards. In this way we want to counteract any possible suspicion that the random event could be manipulated. The subjects see the playing cards and can be sure that when they choose the algorithm there is a probability of 70% that they will be successful (the pack of cards consists of seven +€4 cards and three ±€0 cards). In addition, they can be sure that if they choose a human expert their probability of success is 60% (the pack of cards consists of six +€4 cards and four €±0 cards) (see Appendix D, Figure A-1 and A-2).

The time needed for reading the instructions of the game (Appendix A), answering the test questions (Appendix B) and carrying out the task is 10 minutes on average. A show-up fee of €2 and the possibility of a performance-related payment of €4 seem appropriate for the time spent - it is intended to be sufficient incentive for meaningful economic decisions, and the subjects do actually give the impression of being concentrated and motivated.

The results of the manipulation check show that the subjects perceive the gravity of the decision-making situations significantly differently (Table 3 and Figure 2). In the decision-making situations with serious consequences (Treatment A), the average of the perceived gravity is 9.0 with a

standard deviation of 1.37. In the evaluation of the gravity of the decision-making situations of Treatment B on the other hand, there is an average of 6.54 with a standard deviation of 2.53.

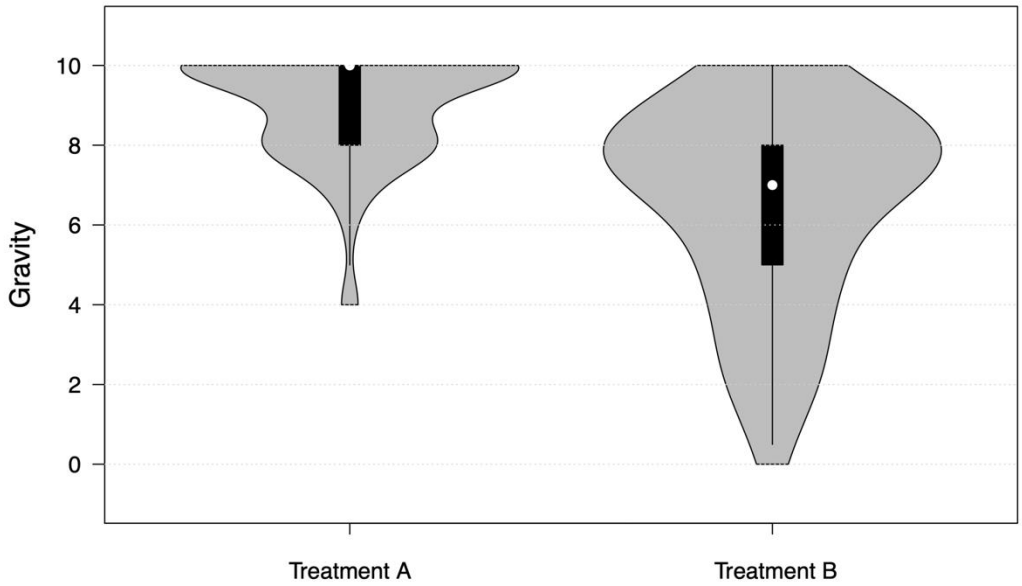
Table 3: Evaluation of gravity in Treatments A and B

	Treatment A	Treatment B
First quartile	8	5
Third quartile	10	8
Median	10	7
Average	9.00	6.54
Standard deviation	1.37	2.53

The graphical analysis also shows that overall, the subjects assess the gravity in Treatment A to be more serious than in Treatment B. In a direct comparison of the violin plots, however, it is obvious that there is a larger range than in Treatment A, because some subjects also assess the gravity as very high in Treatment B (Figure 2).

The Wilcoxon rank-sum test (Mann-Whitney U test) (cf. Wilcoxon, 1945; Mann & Whitney, 1947) shows that the gravity of the decision-making situations in Treatment A is assessed as being significantly higher than that of the decision-making situations in Treatment B ($z = 6.689$; $p \leq 0.001$).

Figure 2: Violin plot for the assessment of the gravity of the decision-making situations



Overall, only 87 out of 143 subjects (60.84%) decide to delegate the service to the (superior) algorithm. A total of 56 subjects (39.16%) prefer to rely on human experts in spite of the lower probability of success. Null hypothesis 1 thus has to be rejected. The result of the Chi-square goodness of fit test is highly significant (χ^2 (n = 143) = 21.93, $p \leq 0.001$). On average, around two

out of five subjects thus tend towards algorithm aversion (Table 4). This is a surprisingly clear result, as the decision-making situations are very obvious. The fact that preferring human experts and rejecting the algorithm reduces the expected value of the performance-related payment should really be completely clear to all of the subjects. However, the need to decide against the algorithm is obviously strong in a part of the subjects.

Table 4: Decisions for and against the algorithm

			Decisions for the algorithm		Decisions against the algorithm	
		n	Number	Percent	Number	Percent
Treatment A	(serious)	71	36	50.70%	35	49.30%
Treatment B	(not serious)	72	51	70.83%	21	29.17%
Σ		143	87	60.84%	56	39.16%

Furthermore, a difference in the number of decisions in favor of the algorithm between the two treatments can be observed. Whereas in Treatment A 50.7% of the subjects trust the algorithm, this figure rises to 70.83% in Treatment B. Karl Pearson's χ^2 test (cf. Pearson, 1900) reveals that null hypothesis 2 has to be rejected ($p = 0.014$). The frequency with which algorithm aversion occurs is influenced by the implications involved in the decision-making situation. The framing effect has an impact.

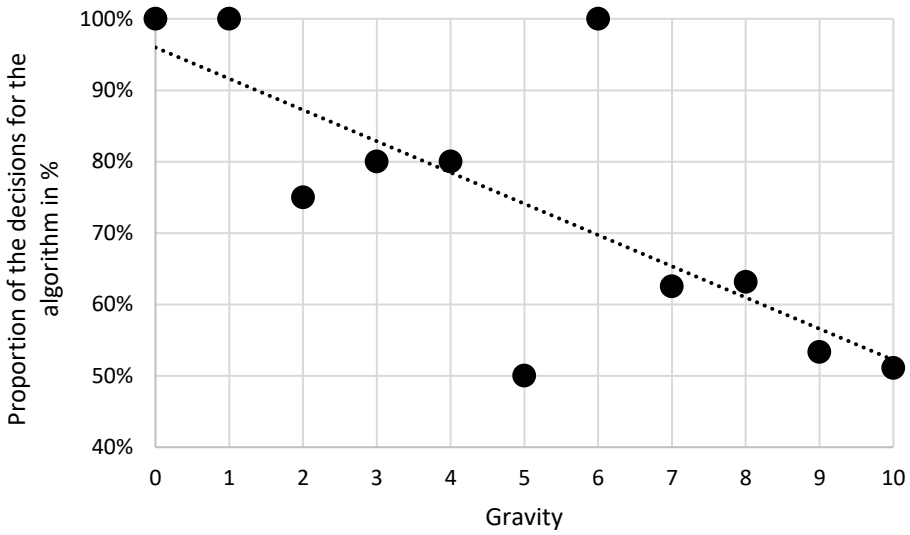
A framing effect sets in, but not in the way one might expect. Whereas in Treatment A (possibly serious consequences) 49.3% of the subjects do exhibit the behavioral anomaly of algorithm aversion, this is only the case in 29.17% of the subjects in Treatment B (no serious consequences) (Table 4). Null hypothesis 3 can therefore not be rejected.

There may be situations in which people like to act irrationally at times. However, common sense suggests that one should allow oneself such lapses in situations where serious consequences must not be feared. For example, when there is a nice barbecue going on, the host may open a third barrel of beer although he suspects that this will lead to hangovers the next day among some of his guests. In the case of important decisions, however, one should be wide awake and try to distance oneself from reckless tendencies. For example, if the same man visits a friend in hospital whose life would be acutely threatened by drinking alcohol after undergoing a complicated stomach operation, he would be wise to avoid bringing him a bottle of his favorite whisky. This comparison of two examples illustrates what could be described as common sense and would be approved of by most neutral observers.

Nevertheless, the results of the experiment point in the opposite direction. In the less serious decision-making situations (Treatment B) the tendency towards algorithm aversion is much less marked than in the serious situations (Treatment A).

This result is confirmed by a regression analysis which demonstrates the relationship between algorithm aversion and the perceived gravity of the decision-making situation. For the possible assessments of the consequences (from 0 = not serious to 10 = very serious), the respective average percentage of the decisions in favor of the algorithm is determined. The decisions of all 143 subjects are included in the regression analysis. Differentiation between the two treatments does not play a role here (Figure 3).

Figure 3: Decisions in favor of the algorithm depending on the gravity of the decision-making situation



If the common sense described above would have an effect, the percentage of decisions for the algorithm from left to right (in other words with increasing perceived gravity of the decision-making situation) would tend to rise. Instead, the opposite can be observed. Whereas in the case of only a low level of gravity (zero and one) 100% of decisions are still made in favor of the algorithm, the proportion of decisions for the algorithm decreases with increasing gravity. In the case of very serious implications (nine and ten), only somewhat more than half of the subjects decide to have the service carried out by an algorithm (Figure 3). If the perceived gravity of a decision increases by a unit, the probability of a decision in favor of the algorithm falls by 3.9% (t: -2.29; p = 0.023). Null hypothesis 3 can therefore not be rejected. In situations which have serious consequences in the case of an error, algorithm aversion is actually especially pronounced.

These results are very surprising, given that common sense would deem – particularly in the case of decisions which have serious consequences – that the option with the greatest probability of success should be chosen. If subjects allow themselves to be influenced by algorithm aversion to make decisions to their own detriment, they should only do so when they can take responsibility for the consequences with a clear conscience. In cases where the consequences are particularly severe, maximization of the success rate should take priority. But the exact opposite is the case. Algorithm aversion appears most frequently in cases where it can cause the most damage. To this extent it seems necessary to speak of the tragedy of algorithm aversion.

The decisive advantage of a framing approach is that the influence of a factor can be clearly identified. There is only one difference between the decision-making situations in Treatment A and Treatment B: the gravity of the possible consequences. It is needless to say that these are consequences which might have to be borne by third parties. It would be possible to continue this line of research by giving up the framing approach and modeling a situation where the subjects are directly affected. In this case, different incentives would have to be introduced into the two treatments. Success in Treatment A (possible serious consequences) would then have to be rewarded with a higher amount than in Treatment B (no serious consequences). However, we presume that our results would also be fully confirmed by an experiment based on this approach, given that it is a between-subjects design in which every subject is only presented with one of the six decision-making situations. Whether one receives €4 or €8 for a successful choice in Treatment A will probably not have a notable influence on the results. Nonetheless, the empirical examination of this assessment is something which will have to wait for future research efforts.

4. Summary

Many people decide against the use of an algorithm even when it is clear that the algorithm promises a higher probability of success than a human mind. This behavioral anomaly is referred to as algorithm aversion.

The subjects are placed in the position of a businessperson who has to choose whether to have a service carried out by an algorithm or by a human expert. If the service is carried out successfully, the subject receives a performance-related payment. The subjects are informed that using the respective algorithm leads to success in 70% of all cases, while the human expert is only successful in 60% of all cases. In view of the recognizably higher success rate, there is every reason to trust in the algorithm. Nevertheless, just under 40% of the subjects decide to use the human expert and not the algorithm. In this way they reduce the expected value of their performance-related payment and thus manifest the behavioral anomaly of algorithm aversion.

The most important objective of the study is to find out whether decision-making situations of varying gravity can lead to differing frequencies of the occurrence of algorithm aversion. To do this, we choose a framing approach. Six decision-making situations (three of which have potentially serious effects and another three which could have not very serious consequences) have an identical payment structure. The differing consequences of the decision-making situations do not affect the subjects themselves, but possibly have implications for third parties. Against this background there is no incentive or reason to act differently in each of the six decision-making situations. It is a between-subjects approach – this means that each subject is only presented with one of the six decision-making situations.

The results are clear. In the three decision-making situations with potentially serious consequences for third parties (Treatment A), just under 50% of the subjects exhibit algorithm aversion. In the three decision-making situations with not very serious consequences for third parties (Treatment B), however, less than 30% of the subjects exhibit algorithm aversion.

This is a really surprising result. If a framing effect were to occur, it would have been expected to be in the opposite direction. In cases with implications for freedom or even danger to life (Treatment A), one should tend to select the algorithm as the option with a better success rate. Instead, algorithm aversion shows itself particularly strongly here. If it is only a matter of arranging a date, creating a weather forecast or offering recipes (Treatment B), the possible consequences are quite clear. In a situation of this kind, one can still afford to have irrational reservations about an algorithm. Surprisingly, however, algorithm aversion occurs relatively infrequently in these situations.

One can call it the tragedy of algorithm aversion because it arises above all in situations in which it can cause particularly serious damage.

References

Alexander, V., Blinder, C. & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology, *Computers in Human Behavior*, *89*(2018), 279-288.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn, *Business & Information Systems Engineering, 63*(1), 55-68.

Brozovsky, L. & Petříček, V. (2007). Recommender System for Online Dating Service, *Proceedings of Znalosti 2007 Conference*, VSB, Ostrava.

Burton, J., Stein, M. & Jensen, T. (2020). A Systematic Review of Algorithm Aversion in Augmented Decision Making, *Journal of Behavioral Decision Making*, *33*(2), 220-239.

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

Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence, *Journal of Accounting Research, 60*(1), 171-201.

Cornelissen, J. & Werner, M. D. (2014). Putting Framing in Perspective: A Review of Framing and Frame Analysis across the Management and Organizational Literature, *The Academy of Management Annals*, *8*(1), 181-235.

Dawes, R., Faust, D. & Meehl, P. (1989). Clinical Versus Actuarial Judgment, *Science*, *243*(4899), 1668-1674.

De-Arteaga, M., Fogliato, R., & Chouldechova, A. (2020). A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Paper 509, 1-12.

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.

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*(C), 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., Judek, J. R., Lorenz, M. & Spiwoks, M. (2021). Reducing Algorithm Aversion through Experience, *Journal of Behavioral and Experimental Finance*, *31*, 100524.

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

Frey, B. S. (1992). Economics As a Science of Human Behaviour, Kluwer Publishing, Dordrecht.

Germann, M., & Merkle, C. (2020). Algorithm Aversion in Financial Investing, SSRN Working Paper, https://dx.doi.org/10.2139/ssrn.3364850 (accessed on August 10th, 2022).

Horne, B. D., Nevo, D., O'Donovan, J., Cho, J. H., & Adalı, S. (2019). Rating Reliability and Bias in News Articles: Does AI Assistance Help Everyone?, *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 247-256.

Ireland, L. (2020). Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors, *Journal of Crime and Justice*, *43*(2), 174-192.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion, *Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1-16.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk, *Econometrica, 47*(2), 263-291.

Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business, *Management Science, 67*(3), 1670-1695.

Köbis, N. & Mossink, L. D. (2021). 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.

Leyer, M., & Schneider, S. (2019). Me, You or Ai? How Do We Feel About Delegation, *Proceedings of the 27th European Conference on Information Systems (ECIS)*, 1-17.

Logg, J., Minson, J. & Moore, D. (2019). Algorithm appreciation: People prefer algorithmic to human judgment, *Organizational Behavior and Human Decision Processes*, *151*(C), 90-103.

Majumdar, A. & Ward, R. (2011). An algorithm for sparse MRI reconstruction by Schatten p-norm minimization, *Magnetic Resonance Imaging*, *29*(3), 408-417.

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other, *Annals of Mathematical Statistics*, *18*(1), 50-60.

Niszczota, P. & Kaszás, D. (2020). Robo-investment aversion, *PLOS ONE*, *15*(9), 1-19.

Önkal, D., Gönül, M. S., & De Baets, S. (2019). Trusting forecasts, *Futures & Foresight Science*, *1*(3-4), 1-10.

Pearson, K. (1900). On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it Can be Reasonably Supposed to have Arisen from Random Sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

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

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

Sawaitul, S. D., Wagh, K. & Chatur, P. N. (2012). Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach, *International Journal of Emerging Technology and Advanced Engineering*, *2*(1), 110-113.

Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles, *Nature Human Behaviour*, *1*(10), 694-696.

Simon, H. A. (1959). Theories of Decision-Making in Economics and Behavioral Science, *The American Economic Review*, 49(3), 253-283.

Simpson, B. (2016). Algorithms or advocacy: does the legal profession have a future in a digital world?, *Information & Communications Technology Law*, *25*(1), 50-61.

Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice, *Science*, *211*(4481), 453-458.

Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases, *Science*, *185*(4157), 1124-1131.

Ueda, M., Takahata, M. & Nakajima, S. (2011). User's food preference extraction for personalized cooking recipe recommendation, *Proceedings of the Second International Conference on Semantic Personalized Information Management: Retrieval and Recommendation*, *781*, 98-105.

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Paper 684, 1-14.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods, *Biometrics Bulletin*, *1*(6), 80-83.

Yeomans, M., Shah, A. K., Mullainathan, S. & Kleinberg, J. (2019). Making Sense of Recommendations, *Journal of Behavioral Decision Making*, *32*(4), 403-414.

Appendix A: Instructions for the game

The game

You are a businessperson and have to decide whether you want a service you are offering for the first time carried out solely by an algorithm or solely by human experts. You are aware that the human experts carry out the task with a probability of success of 60%. You are also aware that the algorithm carries out the task with a probability of success of 70%.

Procedure

After reading the instructions and answering the test questions the decision-making situation is presented to you. This specifies the service which your company offers. First of all, you are asked to assess the gravity of the decision-making situation from the perspective of your customers. Then you decide whether the service should be carried out by human experts or by an algorithm.

Payment

You receive a show-up fee of €2 for taking part in the experiment. Apart from this, an additional payment of €4 is made if the service is carried out successfully.

Information

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

Appendix B: Test questions

Test question 1: Which alternatives are available to you to carry out the service?

a) I can provide the service myself or have it done by an algorithm.
b) I can provide the service myself or have it done by human experts.
c) I can have the service carried out via human experts or by an algorithm. (correct)

Test question 2: For how many newly-offered services do you need to make a choice?

a) None.
b) One. (correct)
c) Two.

Test question 3: How much is the bonus payment for carrying out the task successfully?

a) €1.
b) €2.50.
c) €4. (correct)

Test question 4: How much is the bonus payment if you carry out the task wrongly?

a) -€2.50.
b) €0. (correct)
c) €2.50.

Appendix C: Decision-making situations in Treatments A and B

Appendix C.1: Treatment A - Rather serious decision-making situations

Decision-making situation A-1: Autonomous driving

You are the manager of a public transport company and have to decide whether you want to transport your 100,000 passengers solely with autonomous vehicles (algorithm) or solely with vehicles with drivers (human experts). The task will be considered to have been successfully completed when all of your customers have reached their destination safely. In an extreme case, a wrong decision could mean the death of a passenger.

I choose: O Autonomous vehicles (algorithm)

O Drivers (human experts)

Decision-making situation A-2: MRI scan

You are the manager of a large hospital and have to decide whether the MRI scans of your 100,000 patients with brain conditions should be assessed solely by a specialized computer program (algorithm) or solely by doctors (human experts). The task will be considered to have been successfully completed when all life-threatening symptoms are recognized immediately. In an extreme case, a wrong decision could mean the death of a patient.

I choose: O Specialized computer program (algorithm)

O Doctors (human experts)

Decision-making situation A-3: Criminal cases

You are the head of a large law firm and have to decide whether the analysis of the case documents of your 100,000 clients should be carried out exclusively by a specialized computer program (algorithm) or solely by defense lawyers (human experts). The task will be considered to have been successfully completed when the penalties issued to your clients are below the national average. In an extreme case, a wrong decision could mean an unjustified long prison sentence for a client.

I choose: O Specialized computer program (algorithm)

O Defense lawyers (human experts)

Appendix C.2: Treatment B - Less serious decision-making situations

Decision-making situation B-1: Dating service

You are the manager of an online dating site and have to decide whether potential partners are suggested to your 100,000 customers solely by a specialized computer program (algorithm) or exclusively by trained staff (human experts). The task will be considered to have been successfully completed when you can improve the rating of your app in the App Store. For your customers, a wrong decision could lead to a date with a sub-optimal candidate.

I choose: O Specialized computer program (algorithm)

O Trained staff (human experts)

Decision-making situation B-2: Recipes

You are the manager of an online food retailer and have to decide whether your 100,000 cooking boxes – with ingredients and recipes which are individually tailored to the customers – are put together solely by a specialized computer program (algorithm) or solely by trained staff (human experts). The task will be considered to have been successfully completed when you can increase the reorder rate as a key indicator of customer satisfaction. A wrong decision could mean that the customers don't like their meal.

I choose: O Specialized computer program (algorithm)

O Trained staff (human experts)

Decision-making situation B-3: Weather forecasts

You are the manager of a news site and have to decide whether your 100,000 daily weather forecasts for various cities are carried out solely by a specialized computer program (algorithm) or exclusively by experienced meteorologists (human experts). The task will be considered to have been successfully completed when the temperatures forecast the previous day do not diverge by more than 1 degree Celsius from the actual temperature. A wrong decision could mean that the readers of the forecasts do not dress suitably for the weather.

I choose: O Specialized computer program (algorithm)

O Experienced meteorologists (human experts)

Appendix D: Determination of the random event with the aid of a lottery

Figure A-1: Pack of cards in the selection of the algorithm

Pack of cards in the selection of the algorithm: seven cards with the event +€4 and three cards with the event €±0.

Figure A-2: Pack of cards in the selection of the human expert

Pack of cards in the selection of the human expert: six cards with the event +€4 and four cards with the event €±0.

Chapter IV

Comparing Different Kinds of Influence on an Algorithm in Its Forecasting Process and Their Impact on Algorithm Aversion

Co-authored by Zulia Gubaydullina, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Submitted to:
Businesses

Published in:
Sofia Diskussionsbeiträge, Sonderforschungsgruppe Institutionenanalyse, No. 21-6, Darmstadt, June 2021. (Discussion Paper in German)
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 21-04, Wolfsburg, June 2021. (Working Paper)

Abstract

Although algorithms make more accurate forecasts than humans in many applications, decision-makers often refuse to resort to their use. In an economic experiment, we examine whether the extent of this phenomenon known as algorithm aversion can be reduced by granting decision-makers the possibility to exert an influence on the design of the algorithm (an influence on the algorithmic input). In addition, we replicate the study carried out by Dietvorst, Simmons & Massey (2018). This shows that algorithm aversion recedes significantly if the subjects can subsequently change the results of the algorithm – and even if this is only by a few percent (an influence on the algorithmic output). The present study confirms that algorithm aversion is reduced significantly when there is such a possibility to influence the algorithmic output. However, exerting an influence on the algorithmic input seems to have only a limited ability to reduce algorithm aversion. A limited opportunity to modify the algorithmic output thus reduces algorithm aversion more effectively than having the ability to influence the algorithmic input.

Keywords

Algorithm aversion, technology adoption, human in the loop, human-computer interaction, experiments, behavioral economics.

JEL Classification

D81, D91, G17, G41, O33.

1. Introduction

Businesses throughout the world are driving the digital transformation. Progress in the field of artificial intelligence (AI) has wide-ranging effects on our everyday lives and is bringing about fundamental changes in all fields of human life (Wamba-Taguimdje et al., 2020; Livingston & Risse, 2019; Makridakis, 2017). Technological progress is also leading to algorithms or algorithmic decision-making systems (ADM systems) being increasingly deployed in a wide range of areas and becoming part of day-to-day life. AI-based algorithms make a considerable contribution towards tasks being completed faster and above all more cheaply (Upadhyay & Khandelwal, 2018). In addition, algorithms can better the performance of humans (from lay persons to experts) in a multitude of areas and make more accurate predictions, including the following examples: forecasts on the performance of employees (Highhouse, 2008), the likelihood of ex-prisoners re-offending (Wormith & Goldstone, 1984), or in making medical diagnoses (Beck et al., 2011; Gladwell, 2007; Grove et al., 2000; Dawes, Faust & Meehl, 1989; Adams et al., 1986).

Nevertheless, in certain fields there is a lack of acceptance for the actual use of algorithms because subjects have reservations about them. This phenomenon, which is known as algorithm aversion, refers to the lack of trust in algorithms which arises in subjects as soon as they recognize that the algorithms sometimes make inaccurate predictions (Jussupow, Benbasat & Heinzl, 2020; Prahl & Van Swol, 2017; Dietvorst, Simmons & Massey, 2015). We therefore focus on the issue of how algorithm aversion can be reduced and how the level of acceptance of algorithms can be increased.

In their study, Dietvorst, Simmons and Massey (2018) reveal a way in which algorithm aversion can be significantly reduced. In their experiment, the subjects can either choose an algorithm or make their own forecasts. Some of the subjects are – if they choose to use an algorithm – allowed to subsequently change the preliminary forecast of the algorithm by a few percentage points up or down (we describe this in our study as an opportunity to influence the 'algorithmic output'). When they have this opportunity to make retrospective changes to the forecasts, significantly more subjects are prepared to consult the algorithm for their forecasts than otherwise.

As long as the subjects are able to change the results of the algorithm (i.e., they have an influence on the algorithmic output), algorithm aversion can be significantly reduced. Decisions in favor of an algorithm are made more frequently if the users retain an element of control over it, whereby the extent to which they are able to modify the algorithm is irrelevant. Furthermore, users who can make slight modifications report that they are no less content with the forecasting process than users who can make unlimited changes. To sum up, users will deploy algorithms more often when they have the final say in how they deal with them (Dietvorst, Simmons & Massey, 2018). So is it crucial for lowering algorithm aversion that users are given an opportunity to influence the algorithmic output, or can algorithm aversion be generally reduced by providing a way of influencing the forecasting process?

Human decision-makers want to influence algorithms instead of being at the mercy of their calculations (Honeycutt, Nourani & Ragan, 2020; Stumpf et al., 2008). In other words, decision-makers need partial control over an algorithm in order to make a decision in favor of its use. Having real or at least perceived control over the decisions to be made satisfies the psychological

needs and personal interests of users (Colarelli & Thompson, 2008). This feeling of control can arise either via a real understanding of the efficiency of an algorithm, or via adaptations to the algorithmic decision-making process which have little or no influence on the functioning or level of performance of an algorithm (Burton, Stein & Jensen, 2020). In other words, if a user is granted control over decisions, this leads to a higher level of acceptance: if a recommendation algorithm for hotel rooms is used which only recommends hotel rooms based on the person's previous search and purchasing behavior, the offers made are less readily accepted. However, if less than ideal offers are included, levels of acceptance of the algorithm improve (Taylor, 2017). Participation in the decision-making process, or a belief that one can influence the decision-making process, can contribute towards the user exhibiting greater trust in a decision (Landsbergen et al., 1997).

In our study, we therefore grant the subjects the opportunity to participate in the design of an algorithm by giving them an influence on the algorithmic input, although we keep the extent of their intervention in the algorithmic input small. In this way the algorithm can almost reach its maximum level of performance; however, this minor intervention could be of great significance in overcoming algorithm aversion (cf. Burton, Stein & Jensen, 2020). We examine whether the opportunity to participate in the design of the algorithm has an effect on its acceptance. In addition, we observe whether influencing the algorithmic input can contribute towards a reduction of algorithm aversion in the same way as influencing the algorithmic output does.

2. Experimental design and hypotheses

In this study, the subjects are asked to forecast the exact price of a share in ten consecutive periods (Appendix A). Here, the price of the share is always the result of four influencing factors (A, B, C and D) which are supplemented by a random influence (Ɛ) (see Filiz et al. 2021; Filiz, Nahmer & Spiwoks, 2019; Meub et al., 2015; Becker, Leitner & Leopold-Wildburger, 2009). First of all, the subjects are familiarized with the scenario and are informed that the influencing factors A, C and D have a positive effect on the share price. This means that - other things being equal - when these influencing factors rise the share price will also rise. The influencing factor B, on the other hand, has a negative effect on the share price. This means that - other things being equal - when the influencing factor B rises, the share price will fall (Table 1). In addition, the subjects are informed that the random influence (Ɛ) has an expected value of zero. However, the random influence can lead to larger or smaller deviations from the share price level which the four influencing factors would suggest.

The subjects are informed of the four influencing factors before each of the ten rounds of forecasting. In addition, they always receive a graphic insight into the historical development of the share price, the influencing factors and the random influence in the last ten periods. In this way, the subjects can recognize in a direct comparison how the levels of the four influencing factors have an effect on the share price during the individual rounds of forecasting. Through test questions we ensure that all subjects have understood this (Appendix B).

Table 1: Influencing factors in the formation of the share price

Influencing factor	Influence	Strength of the influence
A	Positive	Strong
B	Negative	Strong
C	Positive	Strong
D	Positive	Medium

The payment structure provides for a fixed show-up fee of €4 and a performance-related element. The level of the performance-related payment is dependent on the precision of the individual share price forecasts, whereby the greater the precision of the forecasts, the higher the payment (Table 2). The subjects can thus obtain a maximum payment of €16 (€4 show-up fee plus €12 performance-related payment from ten rounds of forecasting).

Table 2: Performance-related payment for the forecasts

Deviation of the forecast from the actual share price	Payment for the forecast
€0 ≤ $\lvert K_t - P_t \rvert$ ≤ €5	€1.20
€5 < $\lvert K_t - P_t \rvert$ ≤ €10	€0.90
€10 < $\lvert K_t - P_t \rvert$ ≤ €15	€0.60
€15 < $\lvert K_t - P_t \rvert$ ≤ €20	€0.30
$\lvert K_t - P_t \rvert$ > €20	€0.00

Whereby K_t = share price at the point of time t, P_t = forecast at the point of time t.

In order to help them make the share price forecasts, a forecasting computer (algorithm) is made available to the subjects. The subjects are informed that in the past the share price forecasts of the forecasting computer have achieved a payment of at least €0.60 per forecast in 7 out of 10 cases. The subjects are thus aware of the fact that the algorithm they are using does not function perfectly. In order to make its forecasts, the algorithm uses the information which it has been given on the fundamental influencing factors, the direction and strength of the influence and the random influence (Ɛ) in a way that maximizes the accuracy and thus the expected payoff. In this way, however, it by no means achieves 'perfect' forecasts (for a detailed description of how the algorithm works, see Appendix D). Based on the same information and the historical share prices, the subjects can make their own assessments. They would, however, be wrong to assume that they can outperform the algorithm in this way. Following the suggestions of the algorithm would thus seem to be the more sensible option. Before making their first share price forecast, the subjects make a one-off decision on whether they wish to base their payment for the subsequent ten rounds of forecasting on their own forecasts or on those made by the forecasting computer. Our set-up is oriented towards that used in the study carried out by Dietvorst, Simmons & Massey (2018).

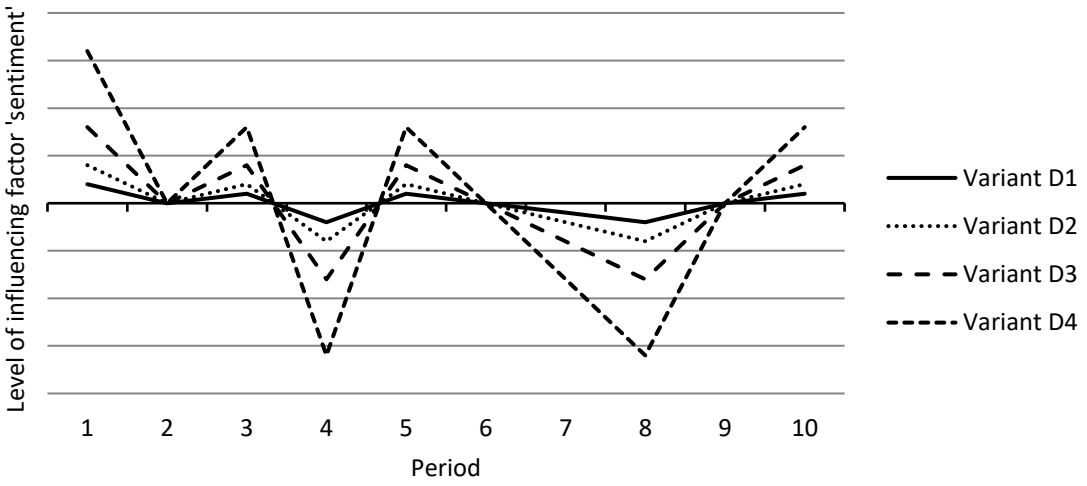
The experiment is carried out in three treatments. The study uses a between-subjects design: each subject is assigned to only one treatment and encounters the respective decision-making situation. In Treatment 1 (no opportunity to influence the algorithm), the subjects make the decision (once only) whether they want to use their own share price forecasts as the basis for their payment or whether they want to use the share price forecasts made by the forecasting computer. Even if the subjects choose the algorithm for determining their bonus, they have to make their own forecasts. In this case, their payoff only depends on the algorithm's forecasts, not on the forecasts made by the subjects themselves. The obligation to submit one's own forecasts even when choosing the algorithm is based on the study by Dietvorst, Simmons & Massey (2018). Regardless of this decision, the subjects make their own forecasts without having access to the forecast of the algorithm. (Figure A-2 in Appendix C).

With Treatment 2 (opportunity to influence the algorithmic output), we intend to replicate the results of Dietvorst, Simmons & Massey (2018). To this end, the subjects make the decision (once only) whether they solely want to use their own share price forecasts as the basis for their payment or whether they solely wish to use the share price forecasts made by the forecasting computer (which, however, can be adjusted by up to +/- €5) as the basis for their performance-related payment. The algorithmic forecast is only made available to the subjects if they decide in favor of the forecasting computer (Figure A-3 in Appendix C).

In Treatment 3, we introduce the opportunity to influence the design of the algorithm (algorithmic input). Before handing in their first share price forecast, the subjects again make the decision (once only) whether they want to solely use their own share price forecasts as the basis for their performance-related payment or whether they want to solely use the share price forecasts made by the forecasting computer. If they decide in favor of the share price forecasts of the forecasting computer, the subjects receive a one-off opportunity to influence the design of the algorithm (Figure A-4 in Appendix C). To this end, they are given a more detailed explanation. The algorithm uses data on four different factors which influence the formation of the share price (A, B, C and D). The last of these four influencing factors is identified as the sentiment of capital market participants and can be taken into account to various extents by the forecasting computer. To do so, the subjects can choose from four different levels. Whereas variant D1 attaches relatively little importance to sentiment, the extent to which sentiment is taken into account in the other variants increases continuously and is relatively strong in variant D4 (Figure 1).

Subjects who decide to use the forecasting computer in Treatment 3 and thus receive the opportunity to influence the design of the algorithm have a one-off chance to change the design of the algorithm. This occurs solely by means of their choice of which degree of sentiment should be taken into account (variant D1, D2, D3 or D4).

Figure 1: Level of the influencing factor 'Sentiment of capital market actors'



Önkal et al. (2009) already point out the phenomenon of algorithm aversion in the field of share price forecasts. Humans rely on share price forecasts less when they have been drawn up by an algorithm instead of a human expert. We examine whether algorithm aversion in the field of share price forecasts also occurs when a choice is made between an algorithm and a subject's own forecasts. The strength of the forecasting computer (algorithm) lies in the fact that it uses the information which it is given in an optimal way. The subjects thus have no reason to expect that they can integrate this information into their forecasts in a similarly efficient way. They should therefore suspect that the forecasting computer will be superior to their own forecasts. On the basis of the existing findings (cf. Jussupow, Benbasat & Heinzl, 2020; Burton, Stein & Jensen, 2020), we expect that algorithm aversion will appear nevertheless. Hypothesis 1 is: The algorithm will not always be selected. Null hypothesis 1 is therefore: All of the subjects will select the algorithm.

Algorithm aversion can be reduced by providing the opportunity to modify the algorithmic output, even when the possibilities for modification are modest (Dietvorst, Simmons & Massey, 2018). In order to establish whether these measures can also contribute towards a reduction of algorithm aversion in the field of share price forecasts, we replicate the above-mentioned results in our experiment. Hypothesis 2 is: The proportion of decisions in favor of the algorithm will be significantly higher in Treatment 2 (opportunity to influence the algorithmic output) than in Treatment 1 (no influence possible). Null hypothesis 2 is therefore: The proportion of decisions in favor of the algorithm will not be significantly higher in Treatment 2 (opportunity to influence the algorithmic output) than in Treatment 1 (no influence possible).

The fact that algorithms can make more accurate predictions than human forecasters has already been shown on numerous occasions (Grove et al., 2000; Dawes, 1979; Meehl, 1954). As modification of the algorithmic output can also have a negative overall effect on forecasting performance, it is examined whether there are additional possibilities to decrease algorithm aversion without allowing human modification of the algorithmic output.

In their review of the literature, Burton, Stein & Jensen (2020) pose the question of whether the reduction of algorithm aversion by the modification of the algorithmic output can also be

achieved by a modification of the algorithmic input. Even the illusion of having the freedom to act and make decisions can be a possible solution to overcome algorithm aversion (Burton, Stein & Jensen, 2020). Users who interact with algorithms often receive their advice from a black box whose workings are a mystery to them. They thus develop theories about which kinds of information an algorithm uses as input and how this information is exactly processed (Logg, Minson & Moore, 2019). According to Colarelli & Thompson (2008), users need to at least have the feeling that they can exercise a degree of control in order to increase the acceptance of algorithms. This feeling of control can either come from a genuine understanding of how an algorithm works or by making modifications to the algorithmic decision-making process. Whether a genuine influence is exerted on the way the algorithm actually functions is not important here. It is only necessary to allow the users to have real or perceived control over decision-making in order to satisfy their need for a feeling of control (Colarelli & Thompson, 2008).

Kawaguchi (2021) has taken a look at how adding an input variable - in this case the predictions made by the subjects - to an algorithm's forecasting process influences algorithm aversion. We draw on this approach and examine how an opportunity to influence the algorithmic input affects the willingness to use the algorithm. In contrast to previous studies, we give our subjects the possibility to influence the weighting of an input factor. In this way we are testing an alternative approach to the reduction of algorithm aversion without influencing the algorithmic output. We do not want to deceive the subjects and thus give them – in the form of this input factor – the opportunity to exert an actual influence on the design of the forecasting computer. In this way, the subjects are given freedom to act in a limited way, which actually leads to slight differences in how the algorithm works. We ask ourselves whether a general possibility to influence the algorithmic process is sufficient in order to reduce algorithm aversion, or whether an opportunity to influence the results themselves is necessary. We thus examine whether an opportunity to influence the design of the algorithm (algorithmic input) can contribute towards a similar decrease in algorithm aversion as the opportunity to influence the algorithmic output. Hypothesis 3 is: The proportion of decisions in favor of the algorithm will be significantly higher in Treatment 3 (opportunity to influence the algorithmic input) than in Treatment 1 (no influence possible). Null hypothesis 3 is therefore: The proportion of decisions in favor of the algorithm will not be significantly higher in Treatment 3 (opportunity to influence the algorithmic input) than in Treatment 1 (no influence possible).

3. Results

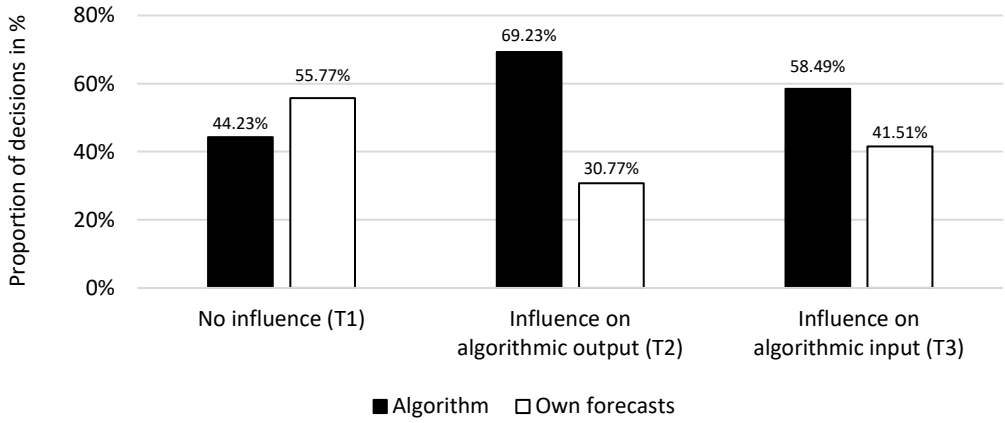
This economic experiment is carried out between 17-27 March 2021 in the Ostfalia Laboratory of Experimental Economic Research (OLEW) with students of the Ostfalia University of Applied Sciences in Wolfsburg. In 51 sessions, a total of 157 subjects take part in the experiment. 118 subjects (75.16%) are male, and 39 subjects (24.84%) are female. The subjects are distributed across the faculties as follows: 66 subjects (42.04%) study at the Faculty of Vehicle Technology, 56 subjects (35.67%) at the Faculty of Business, 9 subjects (5.73%) at the Faculty of Health Care and a further 26 subjects (16.56%) at other faculties based at other locations of the Ostfalia University of Applied Sciences. Their average age is 23.6 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 in line with the measures to contain the Covid-19 pandemic a considerable distance can be maintained between 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 51 sessions with a maximum of four subjects per session are carried out. A session lasts an average of 30 minutes.

The 157 participants are divided up evenly over the three treatments, so that 52 subjects carry out Treatments 1 and 2, and 53 subjects carry out Treatment 3. The distribution of the subjects among the three treatments has similarities to their distribution among the faculties as well as to their gender.

The results show that the various possibilities to influence the forecasting process lead to different decisions on the part of the subjects. In Treatment 1 (no influence possible), 44.23% of the subjects opt for the use of the algorithm. The majority of the subjects here (55.77%) put their faith in their own forecasting abilities. In Treatment 2 (opportunity to influence the algorithmic output) on the other hand, 69.23% of the subjects decide to use the forecasting computer and 30.77% of the subjects choose to use their own forecasts. In Treatment 3 (opportunity to influence the algorithmic input), 58.49% of the subjects decide to use the forecasting computer and 41.51% of the subjects choose to use their own forecasts (Figure 2).

Figure 2: Comparison of the decisions in favor of the algorithm or the subjects' own forecasts per treatment



Overall, 67 subjects (42.68%) decide against using the algorithm. In our study too, the phenomenon of algorithm aversion in the field of share price forecasts shows itself (Önkal et al., 2009; Castelo, Bos & Lehman, 2019). Null hypothesis 1 thus clearly has to be rejected. Slightly more than two fifths of the subjects are affected by the phenomenon of algorithm aversion and

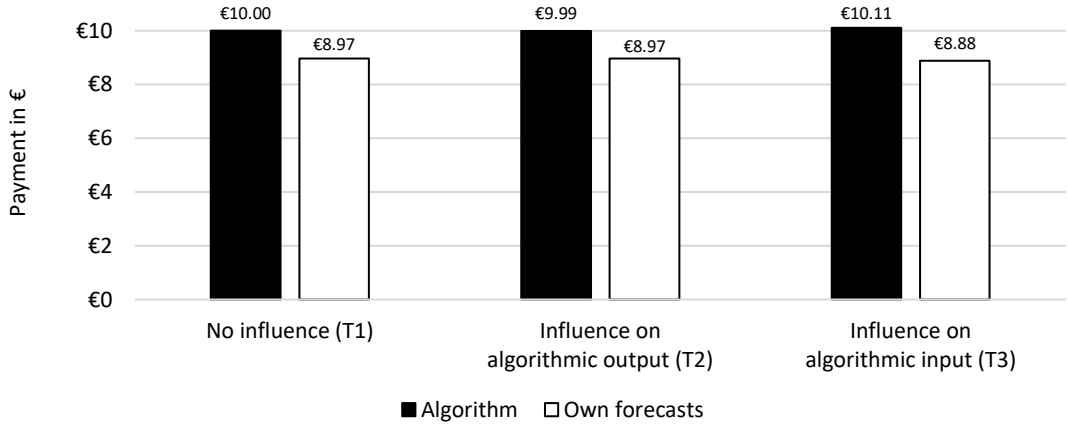
thus miss out on a higher average payment. The significance of this result is confirmed by the Chi-square goodness of fit test (χ^2 (n = 157) = 28.59, p ≤ 0.001).

We subject the distribution of the participants among the individual treatments to the Chi-square test. Whereas in Treatment 1 a total of 44.23% of the decisions are in favor of the algorithm, 69.23% of the subjects who can make changes to the algorithmic output (Treatment 2) decide to use the forecasting computer (χ^2 (n = 104) = 6.62, p ≤ 0.010). Null hypothesis 2 thus has to be rejected; the opportunity to modify the algorithmic output by up to +/- €5 leads to the subjects selecting the forecasting computer significantly more frequently to determine their payment.

When subjects are given the opportunity to influence the algorithmic input (Treatment 3), the majority of the subjects – 58.49 percent – choose to use the forecasting computer (χ^2 (n = 105) = 2.14, p ≤ 0.144). Nevertheless, null hypothesis 3 is not rejected. The possibility to influence the algorithmic input (via the extent to which the influencing factor D is taken into account) does not lead to the subjects selecting the forecasting computer significantly more often as the basis for their performance-related payment.

On average across all three treatments, the subjects obtain a payment of €9.57. However, there are differences in the amounts of the payment depending on the strategy chosen (Figure 3). Subjects who choose their own forecasts achieve an average total payment of €8.94. When the algorithm is chosen, the average payment in all three treatments is between €9.99 and €10.11. The Wilcoxon rank-sum test shows that the average payment – regardless of the treatment – is significantly higher if the algorithm is used as the basis of the forecasts (T1: z = 4.27, p ≤ 0.001; T2: z = 3.25, p ≤ 0.001; T3: z = 5.27, p ≤ 0.001). No matter which treatment is involved, it is thus clearly in the financial interests of the subjects to put their faith in the algorithm.

Figure 3: Average payment in the three treatments depending on the strategy chosen when making the forecasts (own forecast or delegation to the algorithm)



The 67 subjects who, regardless of which treatment they are in, use their own forecasts as the basis of their payment, diverge by an average of €18.28 from the actual share price and thus achieve an average bonus of €0.49 per round of forecasting. Subjects who decide to use the forecasts of the forecasting computer exhibit a lower average forecasting error independently of

which treatment they are in. The average bonus and the average payment of the subjects who use the forecasting computer are also higher than that of subjects who rely on their own forecasting abilities (Table 3).

Table 3: Performance of the subjects in relation to their chosen strategy when making their forecasts (own forecasts or delegation to the algorithm)

	n	Ø Forecast error [€]*	Ø Bonus per round [€]	Ø Total payment [€]
Own forecasts	67	18.2776	0.4939	8.94
Forecasts by the algorithm without the opportunity to influence it *(Treatment 1)*	23	13.4000	0.6000	10.00
Forecasts by the algorithm with an opportunity to influence the output *(Treatment 2)*	36	13.5167	0.5992	9.99
Forecasts by the algorithm with an opportunity to influence the input *(Treatment 3)*	31	13.2968	0.6106	10.11
Total	157	15.4879	0.5566	9.57

***** Ø Deviation between the forecasted share price and the actually occurring share price

In Treatment 2, the subjects are given the opportunity to adapt the algorithmic output in each round of forecasting by up to +/- €5. The subjects do not fully exploit the scope granted to them to exert an influence on the algorithm and make an average change to the algorithmic forecast of €2.11. In Treatment 3 the subjects are given a one-off opportunity via the influencing factor D (sentiment) to exert an influence on the design of the algorithm (input). Eight subjects select variant D1, which takes sentiment into account to a minor extent. Eleven subjects choose to take sentiment into account to a moderate extent, seven to a considerable extent, and five to a great extent.

If the results are viewed in isolation, a similar picture is revealed. Regardless of whether subjects use their own forecasts or the forecasts of the forecasting computer to determine their payment, the average forecast error in Treatment 1 (no influence possible) is higher than in the other two treatments, which offer the subjects the opportunity to influence the algorithm. Whereas the forecasts in Treatment 1 deviate by an average of €16.18 from the resulting share price, the average forecast error in Treatment 2 is €15.14 and €15.15 in Treatment 3. That those subjects who are given the opportunity to influence the algorithm are more successful is shown by their average bonus and higher average overall payment (Table 4).

Table 4: Comparison of the performance of the subjects across all three treatments

	n	Ø Forecast error [€]*	Ø Bonus per round [€]	Ø Total payment [€]
No influence possible *(Treatment 1)*	52	16.1788	0.5423	9.42
Influence on the algorithmic output *(Treatment 2)*	52	15.1442	0.5677	9.68
Influence on the algorithmic input *(Treatment 3)*	53	15.1472	0.5598	9.60
Total	157	15.4879	0.5566	9.57

* Ø Deviation between the forecasted share price and the actually occurring share price

4. Discussion

Algorithm aversion is characterized by the fact that it mostly occurs when algorithms recognizably do not function perfectly (Dietvorst, Simmons & Massey, 2015). Even when it is recognizable that the algorithm provides significantly more reliable results than humans (lay persons as well as experts), many subjects are still reluctant to trust the algorithm (Dietvorst, Simmons & Massey, 2018). In our study too, the forecasts of the computer (algorithm) are far from perfect, and the majority of users choose not to use the forecasting computer if there are no opportunities to influence the programme's decision-making process.

Nevertheless, the results show that subjects who are granted an opportunity to influence the algorithmic input or output are more successful on average than subjects who do not have this opportunity. This is because they make the algorithm into the basis for their payment more frequently, and it can be viewed as the success which comes from a reduction in algorithm aversion. On the other hand, forecasters who trust in their own forecasts not only make less accurate forecasts overall; they also obtain lower payment for their efforts (Table 3 and Table 4). It can also be seen that the forecasts of the forecasting computer after the changes made to the algorithm by the subjects in Treatments 2 and 3 are almost equally successful to those made without this possibility in Treatment 1 (Table 3). However, the subjects in Treatment 1 (no opportunity to influence the algorithm) put their faith in the forecasting computer relatively seldom and thus reduce their bonus and their overall payment.

In Treatment 2 (opportunity to influence the algorithmic output), we replicate the results of Dietvorst, Simmons and Massey (2018). In our study too, the forecasters tend to rely on the forecasting computer significantly more often if they have at least a small opportunity to influence the algorithmic output. Even though the subjects do not fully exploit this opportunity to make modifications, it does lead to them being more successful overall, to them making less errors on average, and to them achieving a higher bonus than subjects who are not given the chance to influence the algorithm (Treatment 1) (Table 4). This is not because they make better forecasts - they simply use the (superior) algorithm more frequently.

In Treatment 3 (opportunity to influence the algorithmic input) we obtain similar results. The possibility of influencing the algorithmic input also seems to be suited to reducing algorithm aversion. Nevertheless, the differences in comparison to Treatment 1 are not statistically significant. It is the chance to exert a minor influence on the algorithmic output which reduces algorithm aversion tremendously (Dietvorst, Simmons & Massey, 2018). We ask ourselves whether this major reduction in algorithm aversion is due to the fact that the subjects can exercise an influence on the process of algorithmic decision-making in general, or only because they can influence its results. Here we can see that a general opportunity to influence the algorithm is obviously not sufficient to significantly reduce algorithm aversion. Subjects want to retain control over the results and to have the final say in the decision-making process, even if this intervention is limited by considerable restrictions.

Nevertheless, our study has interesting implications for real-life situations. The overall financial benefit can be maximized by influencing the algorithmic output. Decision-makers tend to trust an algorithm more if they can keep the upper hand in the decision-making process. This even applies when the possibilities to exert an influence are limited. In our study, the average share price is €100 and the maximum amount by which the results can be adjusted is €5. This corresponds to just five percent of the average value of the subject of the forecast, and yet it suffices to significantly shift the grounds on which the decision is based in favor of the algorithm. The average quality of the forecasts is slightly reduced due to the changes made by the decision-maker (Table 3), but this is over-compensated for by a significantly higher utilization rate of the – still clearly superior – algorithm, and in a comparison between the treatments this leads to a higher average total payment (Table 4). The opportunity to influence the algorithmic input has a similar effect with regard to the overall pecuniary benefit. The forecasts made after the subjects have made changes to the algorithm actually exhibit a slightly lower forecast error and a somewhat higher bonus. To a similar degree to which the subjects do not fully take advantage of the opportunity to influence the algorithmic output, they also fail to put their faith in the algorithm. Their average payment is nevertheless significantly higher than that of the subjects who cannot influence the algorithm.

Our study also has some limitations which should be noted. We give the subjects a genuine opportunity to influence the algorithmic input. However, we also make it clear in the instructions that the influencing factor D, which can be taken into account to different degrees, only has a moderate influence on the formation of the share price. The influencing factors A, B and C, on the other hand, have a considerable influence. This circumstance could contribute towards the subjects not developing enough trust in their opportunity to influence the input and thus tending to rely on their own forecasts.

Future research work may wish to investigate further possibilities to reduce algorithm aversion. This study has again shown that granting subjects the opportunity to influence the algorithmic output can effectively reduce algorithm aversion. However, there is a risk that the forecasting performance of the algorithm can deteriorate as a result of the modifications. For this reason, it is important to examine alternative forms of reducing algorithm aversion. Our study has shown that modifying the algorithmic input is only of limited use here. Opportunities to influence the algorithmic input cannot reduce algorithm aversion to the same extent as giving subjects the chance to influence the algorithmic output. We therefore recommend that further research be

carried out to search for other alternatives to reduce algorithm aversion. One possible approach could be to merely give users the illusion of having control over the algorithmic process. In this way, algorithm aversion could be decreased without a simultaneous reduction of the forecasting quality.

5. Conclusion

In this study we carry out an experiment to investigate in which ways algorithm aversion can be reduced. It is a well-known fact that providing subjects with an opportunity to influence the algorithmic output is a suitable means of significantly reducing algorithm aversion. We examine whether providing a possibility to influence the algorithmic input also contributes towards decreasing algorithm aversion.

In our experiment, the subjects are asked to make forecasts of share prices. In return, they receive a performance-related payment which increases in line with the precision of their share price forecasts. In three treatments the subjects have a forecasting computer (algorithm) available to them whose forecasts deviate by a maximum of €15 from the actual share price in 7 out of 10 cases. In this way they can earn a bonus of at least €0.60 per forecast. The maximum possible payment per forecast is €1.20. The predictions of the forecasting computer are thus by no means perfect. In Treatment 1 we do not grant the subjects any opportunity to influence the forecasting process. In Treatment 2, on the other hand, the subjects are able influence the algorithmic output, and in Treatment 3 they can influence the algorithmic input.

In agreement with the literature on algorithm aversion, we establish that even a considerably limited opportunity to influence the algorithmic output is able to reduce algorithm aversion significantly. However, being able to influence the algorithmic input does not lead to a significant reduction in algorithm aversion. Granting subjects a general possibility to influence the algorithmic decision-making process is therefore not a crucial factor in reducing algorithm aversion. What does lead to a significantly higher rate of using the forecasting computer, however, is the opportunity to influence the algorithmic output. This remains true even when the opportunity to influence the programme is only a minor one. Subjects want to have the upper hand over the algorithm and to have the final say in the decision-making process.

Nevertheless, we note that the overall financial benefit to the subjects can be increased via the opportunity to influence the algorithmic input. Regardless of whether an opportunity to influence the algorithmic input or output is granted, on average the forecasts exhibit similar forecast errors and similar levels of bonuses per round of forecasting. Overall, the subjects achieve a higher payment. If they have the opportunity to influence the algorithmic output, this effect is reproduced even more strongly than in relation to the algorithmic input, given that in the former the proportion of decisions in favor of the algorithm is highest.

References

Adams, I., Chan, M., Clifford, P., Cooke, W. M., Dallos, V., Dombal, F. T., Edwards, M., Hancock, D., Hewett, D. J. & McIntyre, N. (1986). Computer aided diagnosis of acute abdominal pain: a multicentre study, *British Medical Journal, 293*(6550), 800-804.

Beck, A., Sangoi, A., Leung, S., Marinelli, R. J., Nielsen, T., Vijver, M. J., West, R., Rijn, M. V., & Koller, D. (2011). Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival, *Science Translational Medicine*, *3*(108), 108-113.

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

Burton, J., Stein, M. & Jensen, T. (2020). A Systematic Review of Algorithm Aversion in Augmented Decision Making, *Journal of Behavioral Decision Making*, *33*(2), 220-239.

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

Colarelli, S. M. & Thompson, M. B. (2008). Stubborn Reliance on Human Nature in Employee Selection: Statistical Decision Aids Are Evolutionarily Novel, *Industrial and Organizational Psychology*, *1*(3), 347-351.

Dawes, R. (1979). The Robust Beauty of Improper Linear Models in Decision Making, *American Psychologist*, *34*(7), 571-582.

Dawes, R., Faust, D. & Meehl, P. (1989). Clinical Versus Actuarial Judgment, *Science*, *243*(4899), 1668-74.

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, 144*(1), 114-126.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021). Reducing algorithm aversion through experience, *Journal of Behavioral and Experimental Finance*, *31*(5), 100524.

Filiz, I., Nahmer, T. & 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.

Gladwell, M. (2007). Blink: The Power of Thinking Without Thinking, Back Bay Books, New York City.

Grove, W., Zald, D., Lebow, B., Snitz, B. & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis, *Psychological Assessment, 12*(1), 19-30.

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection, *Organizational Psychology*, *1*(3), 333-342.

Honeycutt, D., Nourani, M. & Ragan, E. (2020). Soliciting Human-in-the-Loop User Feedback for Interactive Machine Learning Reduces User Trust and Impressions of Model Accuracy, *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *8*(1), 63-72.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion, *Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1-16.

Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business, *Management Science*, *67*(3), 1670-1695.

Landsbergen, D., Coursey, D. H., Loveless, S. & Shangraw, R. (1997). Decision Quality, Confidence, and Commitment with Expert Systems: An Experimental Study, *Journal of Public Administration Research and Theory*, *7*(1), 131-158.

Livingston, S., & Risse, M. (2019). The Future Impact of Artificial Intelligence on Humans and Human Rights, *Ethics & International Affairs*, *33*(2), 141-158.

Logg, J., Minson, J. & Moore, D. (2019). Algorithm appreciation: People prefer algorithmic to human judgment, *Organizational Behavior and Human Decision Processes*, *151*(C), 90-103.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms, *Futures*, *90*, 46-60.

Meehl, P. (1955). Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence, University of Minnesota Press, Minneapolis.

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

Önkal, D., Goodwin, P., Thomson, M., Gönül, S. & Pollock, A. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments, *Journal of Behavioral Decision Making*, *22*(4), 390-409.

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

Stumpf, S., Sullivan, E., Fitzhenry, E., Oberst, I., Wong, W. K. & Burnett, M. (2008). Integrating rich user feedback into intelligent user interfaces, *Proceedings of the 13th international conference on Intelligent user interfaces*, 50-59.

Taylor, E. L. (2017). Making sense of “algorithm aversion”, *Research World, 2017*(64), 57-57.

Upadhyay, A. K., & Khandelwal, K. (2018). Applying artificial intelligence: implications for recruitment, *Strategic HR Review*, *17*(5), 255-258.

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects, *Business Process Management Journal, 26*(7), 1893-1924.

Wormith, J. S. & Goldstone, C. S. (1984). The Clinical and Statistical Prediction of Recidivism, *Criminal Justice and Behavior*, *11*(1), 3-34.

Appendix A: Instructions for the game

Appendix A.1: Instructions for the game in Treatment 1 (no opportunity to influence the algorithm)

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 10 periods.

The price of share Z is always the result of four influencing factors (**A**, **B**, **C** and **D**) and a random influence (**ε**). The influencing factors are announced before every round of forecasting. In addition, you receive an insight into the past development of the share price, the influencing factors and the random influence in the last ten periods.

The influencing factors **A**, **C** and **D** have a positive effect on the share price. This means that when these influencing factors rise, the share price will also tend to rise (Table A-1).

The influencing factor **B** has a negative effect on the share price. This means that when the influencing factor **B** rises, the share price will tend to fall (Table A-1).

Table A-1: Influencing factors in the formation of the share price (Treatment 1) [1]

Influencing factor	Influence	Strength of the influence
A	Positive	Strong
B	Negative	Strong
C	Positive	Strong
D	Positive	Medium

The random influence **ε** has an expected value of 0, but it can lead to smaller or larger deviations of the share price from the level which the influencing factors would suggest.

You can choose whether your own share price forecasts or the share price forecasts of a forecasting computer (algorithm) are used to determine your payment. Regardless of your choice, you will make your own share price forecasts.

You will receive a show-up fee of €4 for participating. In addition, you receive a performance-related payment: the more accurate your share price forecasts are, the higher your payment. For each forecast made, you receive…

[1] In the original instructions, this table was referred to as "Table 1."

- €1.20 in the case of a deviation of a maximum of €5 of the forecast from the actual share price;
- €0.90 in the case of a deviation of a maximum of €10 of the forecast from the actual share price;
- €0.60 in the case of a deviation of a maximum of €15 of the forecast from the actual share price;
- €0.30 in the case of a deviation of a maximum of €20 of the forecast from the actual share price.

In the past, the share price forecasts of the algorithm have achieved a payment of at least €0.60 per forecast in 7 out of 10 cases.

Procedure

After reading the instructions and answering the test questions, you initially choose whether your own share price forecasts or the forecasts of the forecasting computer (algorithm) are used to determine your payment.

Following this, you will see the price history of share Z, the trend of the influencing factors and the trend of the random influence Ɛ in the last ten periods. In addition, you will receive the influencing factors for the next period. You will be asked to forecast the trend of the share price in the next period.

After making your share price forecast you will see the actual price of share Z. Following this, you will hand in your share price forecasts for the next period. A total of ten rounds are played.

You have a time limit of two minutes available for handing in each share price forecast.

Information

- Please remain quiet during the experiment!
- Please do not look at your neighbor's screen!
- Apart from a pen/pencil and a pocket calculator, **no** other aids are permitted (smartphones, smart watches etc.).
- Only use the sheet of white paper issued to you for your notes.

Appendix A.2: Instructions for the game in Treatment 2 (opportunity to influence the algorithmic output)

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 10 periods.

The price of share Z is always the result of four influencing factors (**A**, **B**, **C** and **D**) and a random influence (**Ɛ**). The influencing factors are announced before every round of forecasting. In addition, you receive an insight into the past development of the share price, the influencing factors and the random influence in the last ten periods.

The influencing factors **A**, **C** and **D** have a positive effect on the share price. This means that when these influencing factors rise, the share price will also tend to rise (Table A-2).

The influencing factor **B** has a negative effect on the share price. This means that when the influencing factor **B** rises, the share price will tend to fall (Table A-2).

Table A-2: Influencing factors in the formation of the share price (Treatment 2) [2]

Influencing factor	Influence	Strength of the influence
A	Positive	Strong
B	Negative	Strong
C	Positive	Strong
D	Positive	Medium

The random influence **Ɛ** has an expected value of 0, but it can lead to smaller or larger deviations of the share price from the level which the influencing factors would suggest.

You can choose the basis which is used to determine your payment:

- Either you can forecast the future share price yourself and forego the use of a forecasting computer (algorithm)
- Or you can use the forecasts of the forecasting computer. <u>If you decide to use the forecasting computer's forecasts (algorithm), you are not bound to the exact forecast provided by the computer. You can change the computer's proposal by up to +/- €5.</u>

[2] In the original instructions, this table was referred to as "Table 1."

You will receive a show-up fee of €4 for participating. In addition, you receive a performance-related payment: the more accurate your share price forecasts are, the higher your payment. For each forecast made, you receive…

- €1.20 in the case of a deviation of a maximum of €5 of the forecast from the actual share price;
- €0.90 in the case of a deviation of a maximum of €10 of the forecast from the actual share price;
- €0.60 in the case of a deviation of a maximum of €15 of the forecast from the actual share price;
- €0.30 in the case of a deviation of a maximum of €20 of the forecast from the actual share price.

In the past, the share price forecasts of the algorithm have achieved a payment of at least €0.60 per forecast in 7 out of 10 cases.

Procedure

After reading the instructions and answering the test questions, you initially choose which basis is used to determine your payment. You can forecast the future share prices without the help of the forecasting computer (algorithm). Or you can use the forecasts of the forecasting computer and change them by up to +/- €5.

Following this, you will see the price history of share Z, the trend of the influencing factors and the trend of the random influence Ɛ in the last ten periods. In addition, you will receive the influencing factors for the next period. You will be asked to forecast the trend of the share price in the next period.

After making your share price forecast you will see the actual price of share Z. Following this, you will hand in your share price forecasts for the next period. A total of ten rounds are played.

You have a time limit of two minutes available for handing in each share price forecast.

Information

- Please remain quiet during the experiment!
- Please do not look at your neighbor's screen!
- Apart from a pen/pencil and a pocket calculator, **no** other aids are permitted (smartphones, smart watches etc.).
- Only use the sheet of white paper issued to you for your notes.

Appendix A.3: Instructions for the game in Treatment 3 (opportunity to influence the algorithmic input)

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 10 periods.

The price of share Z is always the result of four influencing factors (**A**, **B**, **C** and **D**) and a random influence (**Ɛ**). The influencing factors are announced before every round of forecasting. In addition, you receive an insight into the past development of the share price, the influencing factors and the random influence in the last ten periods.

The influencing factors **A**, **C** and **D** have a positive effect on the share price. This means that when these influencing factors rise, the share price will also tend to rise (Table A-3).

The influencing factor **B** has a negative effect on the share price. This means that when the influencing factor **B** rises, the share price will tend to fall (Table A-3).

Table A-3: Influencing factors in the formation of the share price (Treatment 3) [3]

Influencing factor	Influence	Strength of the influence
A	Positive	Strong
B	Negative	Strong
C	Positive	Strong
D	Positive	Medium

The random influence **Ɛ** has an expected value of 0, but it can lead to smaller or larger deviations of the share price from the level which the influencing factors would suggest.

You can choose whether your own share price forecasts or the share price forecasts of a forecasting computer (algorithm) are used to determine your payment. Regardless of your choice, you will make your own share price forecasts.

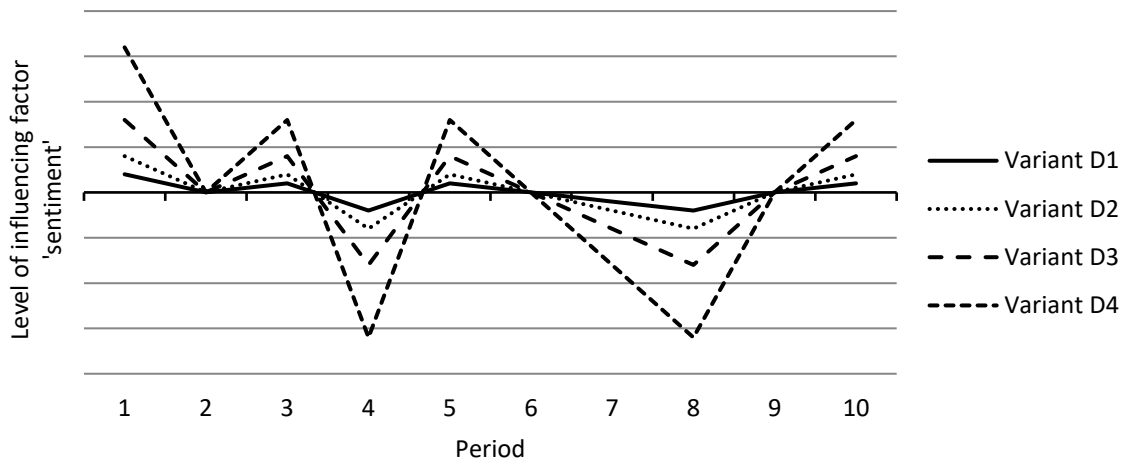
If you decide to use the forecasting computer's forecasts (algorithm), you have the opportunity to influence the design of the algorithm.

As mentioned above, the influencing factor **D** also has an effect on the formation of the price alongside the influencing factors **A**, **B** and **C**. The influencing factor **D** is the sentiment of capital market participants. The influencing factor **D** can be taken into account to differing extents (D1,

[3] In the original instructions, this table was referred to as "Table 1."

D2, D3 or D4) (Figure A-1). You decide which of these four variants should be taken into account by the forecasting computer (algorithm).

Figure A-1: Variants of the influencing factor D (sentiment) [4]



You will receive a show-up fee of €4 for participating. In addition, you receive a performance-related payment: the more accurate your share price forecasts are, the higher your payment. For each forecast made, you receive…

- €1.20 in the case of a deviation of a maximum of €5 of the forecast from the actual share price;
- €0.90 in the case of a deviation of a maximum of €10 of the forecast from the actual share price;
- €0.60 in the case of a deviation of a maximum of €15 of the forecast from the actual share price;
- €0.30 in the case of a deviation of a maximum of €20 of the forecast from the actual share price.

In the past, the share price forecasts of the algorithm have achieved a payment of at least €0.60 per forecast in 7 out of 10 cases.

Procedure

After reading the instructions and answering the test questions, you initially choose whether your own share price forecasts or the forecasts of the forecasting computer (algorithm) are used to determine your payment.

Following this, you will see the price history of share Z, the trend of the influencing factors and the trend of the random influence Ɛ in the last ten periods. In addition, you will receive the

[4] In the original instructions, this figure was referred to as "Figure 1."

influencing factors for the next period. You will be asked to forecast the trend of the share price in the next period.

After making your share price forecast you will see the actual price of share Z. Following this, you will hand in your share price forecasts for the next period. A total of ten rounds are played.

You have a time limit of two minutes available for handing in each share price forecast.

Information

- Please remain quiet during the experiment!
- Please do not look at your neighbor's screen!
- Apart from a pen/pencil and a pocket calculator, **no** other aids are permitted (smartphones, smart watches etc.).
- Only use the sheet of white paper issued to you for your notes.

Appendix B: Test questions

Test question 1: For how many periods should a share price forecast be made?

a) 5.
b) 10. (correct)
c) 15.

Test question 2: On which influences is the share price dependent?

a) Influencing factors A and B as well as the random influence.
b) Influencing factors A, B and C as well as the random influence.
c) Influencing factors A, B, C and D as well as the random influence. (correct)

Test question 3: Which alternatives do you have when submitting your forecast?

a) I can only submit my own forecasts.
b) I can either submit my own forecasts or use a forecasting computer (algorithm). (correct)
c) I can either submit my own forecasts, use a forecasting computer or consult a financial expert.

Test question 4: How much is the payment for a forecast which deviates no more than €15 from the actual price?

a) €1.20.
b) €0.90.
c) €0.60. (correct)

Appendix C: Screens

Figure A-2: Screen when submitting one's own forecasts (Treatments 1, 2 and 3)

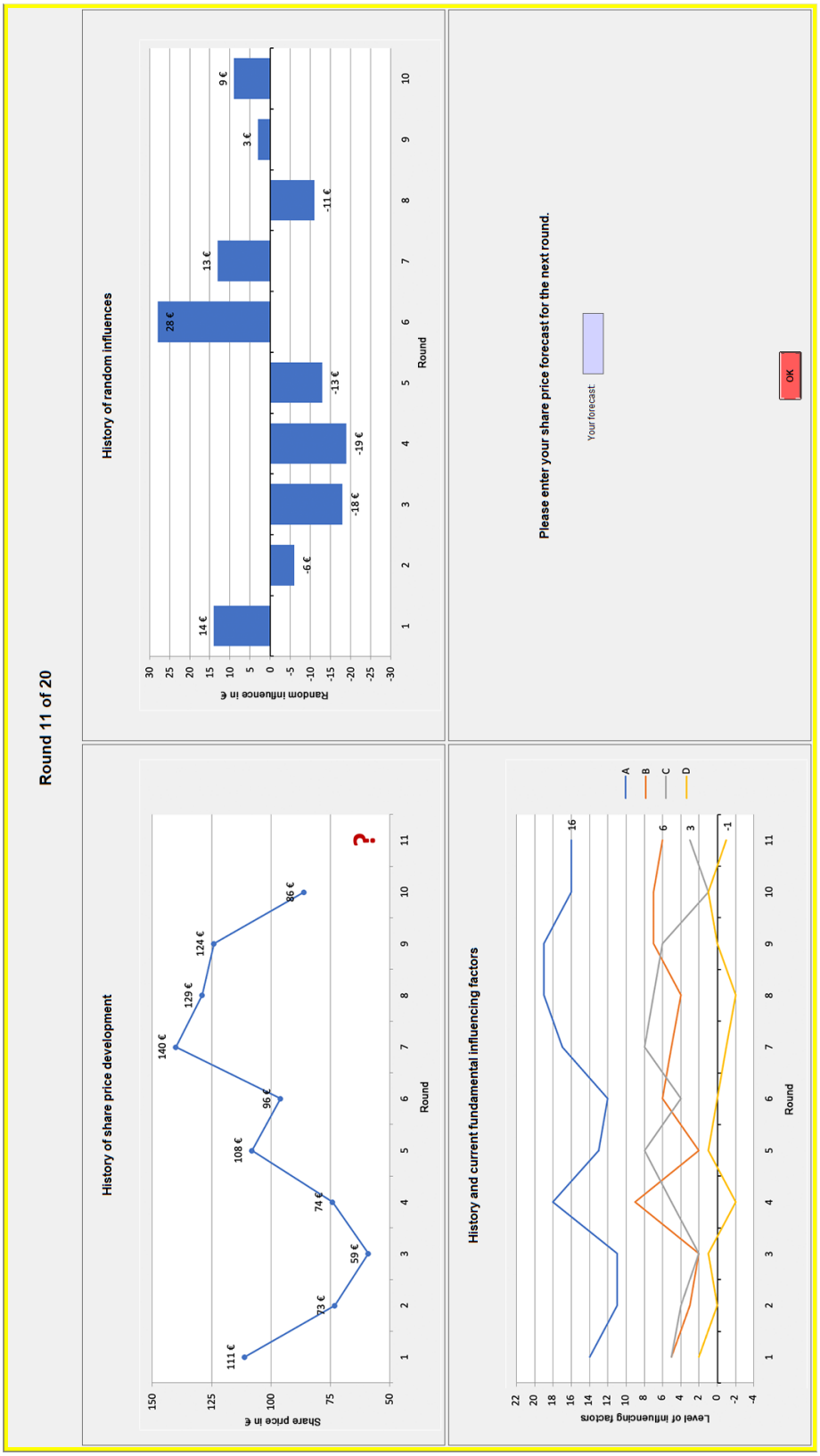


Figure A-3: Screen when influencing the algorithmic forecast (Treatment 2)

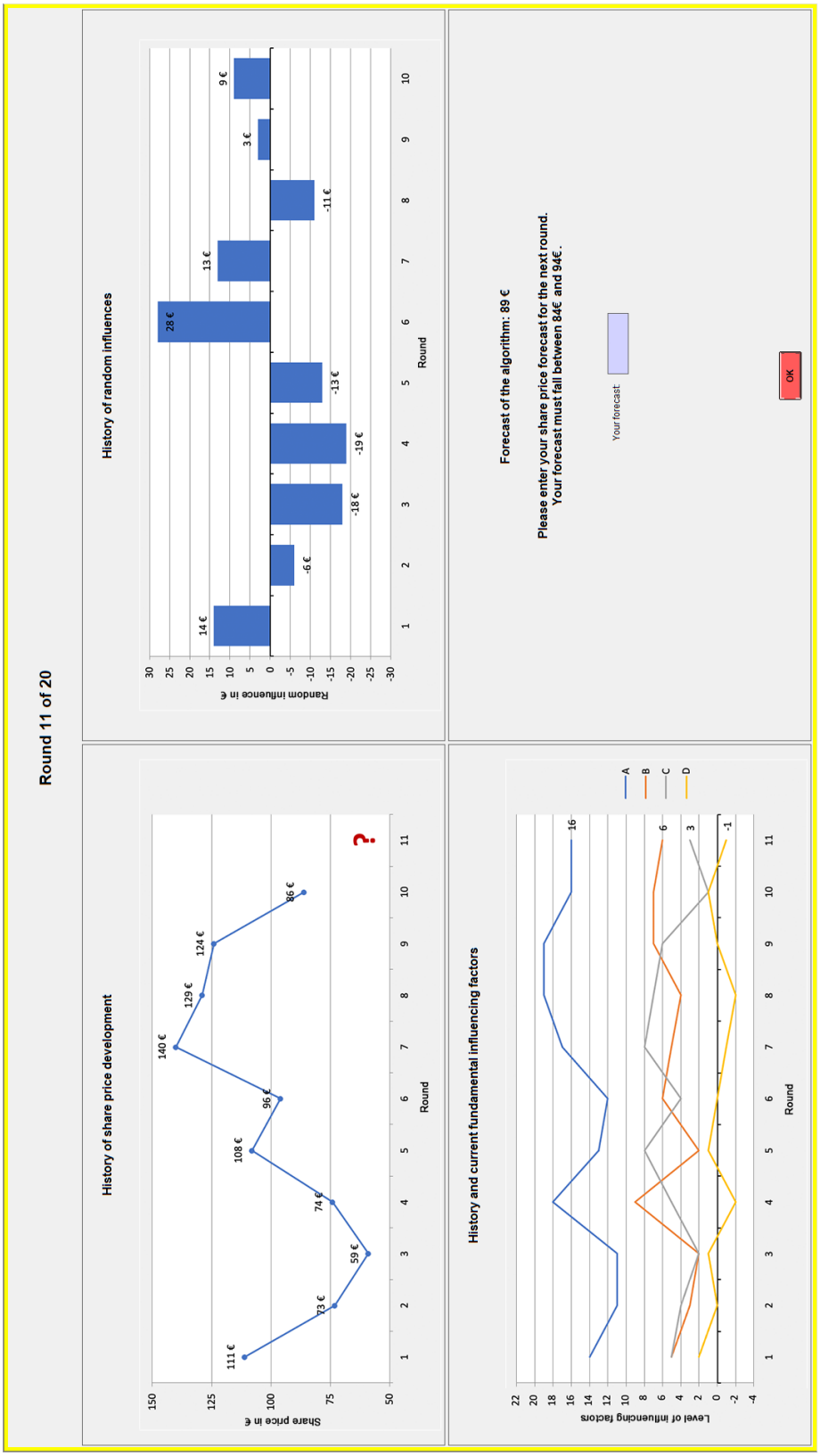
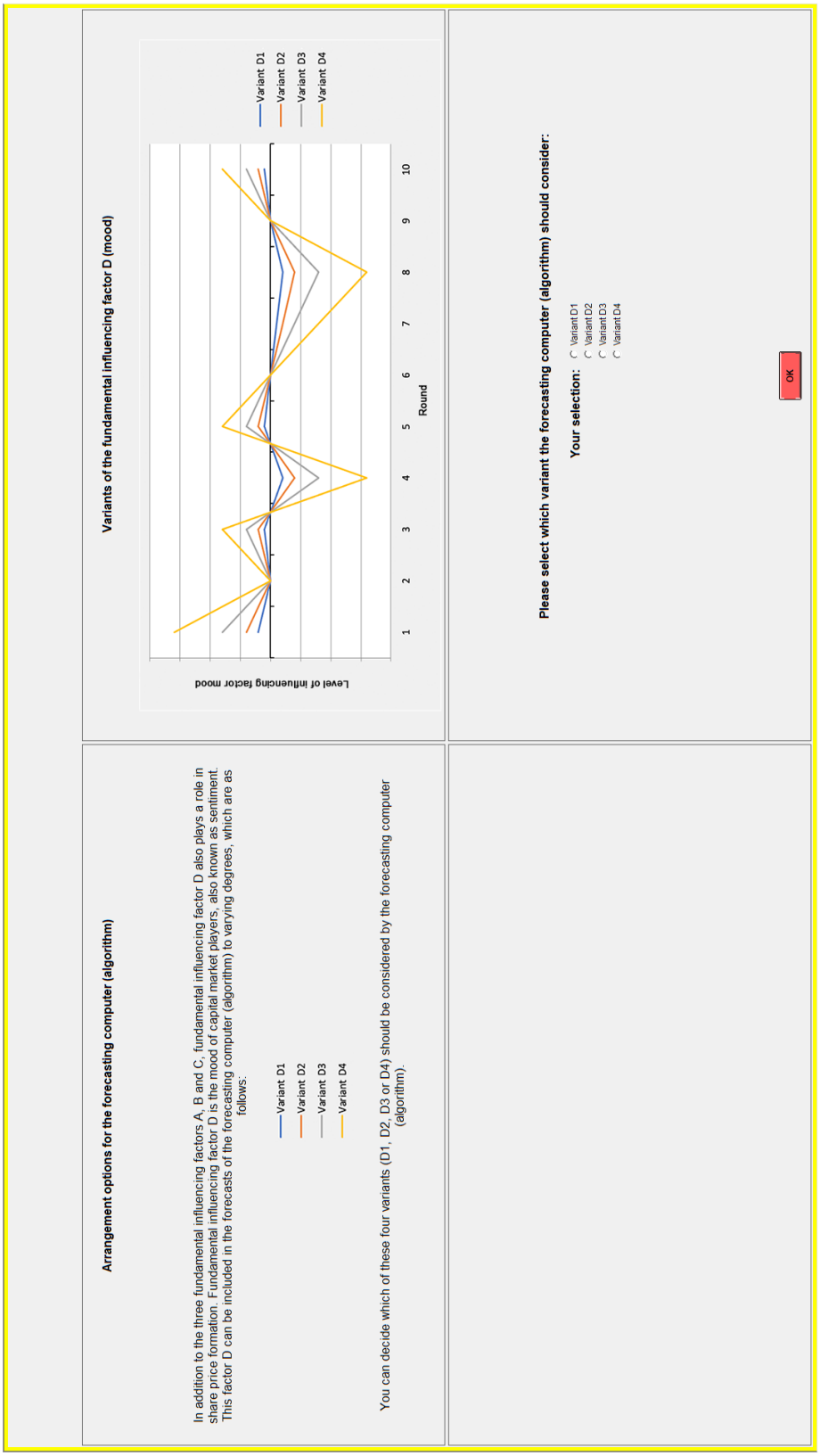


Figure A-4: Screen during the configuration of the algorithm (Treatment 3)



Appendix D: The functioning of the algorithm

The mechanism with which the share price is formed functions as follows:

$$K_t = 7A - 6B + 5C + 2D + \varepsilon$$

The level of the influencing factors A, B, C and D are announced before every round of forecasting. The level of the random influence is not announced. What is known, however, is that the random influence has an expected value of 0. In every round, the algorithm inserts the values of the four influencing factors A, B, C and D into the formula for the formation of the price. Due to the fact that the subjects can influence the algorithmic input, the weighting of the influencing factor D can diverge somewhat in Treatment 3. For the random influence, the algorithm sets the expected value at €0. The result of this equation is the forecast of the algorithm P_t (see Table A-4). In period 1, the algorithm calculates as follows:

$$P_1 = 7 \times 14 - 6 \times 5 + 5 \times 5 + 2 \times 2 + 0 = 97$$

For the calculation of the actual price, the random influence also has an effect. In period 1 it has a value of €+14. The actual price is thus calculated as follows:

$$K_1 = 7 \times 14 - 6 \times 5 + 5 \times 5 + 2 \times 2 + 14 = 111$$

The difference between the actual share price K and the forecast of the algorithm P is the forecast error. This determines the amount of the bonus of the current forecasting round as described in accordance with the formula described in Table 2. For a forecast whose forecast error lies within the interval $10 < |K_t - P_t| \leq 15$ for example, there is a bonus of €0.60.

Table A-4: Illustration of the modus operandi of the algorithm, how the share price is formed, and the calculation of the bonus

Period	Influencing factors				Forecast of algorithm P_t	Random influence	Actual price K_t	Forecast error	Bonus
	A	B	C	D					
1	14	5	5	2	€97	+€14	€111	€14	€0.60

In practice one can see that perfect share price forecasts are not possible, even with knowledge of the most important influencing factors. On the contrary: share price trends have a number of similarities with random processes. This circumstance is taken into account by introducing the random influence. The random influence has the effect that the algorithm cannot make perfect forecasts. The forecast error of the algorithm thus corresponds to the random influence.

In this economic experiment, the random influence consistently lies within the interval -€30 ≤ Ɛ ≤ €30. It is always a whole number without decimal places. The exact distribution is described in Table A-2. The area -€15 ≤ Ɛ ≤ €15 (grey background) has a cumulative probability of 70%. For a forecast with a maximum forecasting error of €15 there is a payment of €0.60. In this way it can be ensured – as stated in the instructions – that the forecasts of the algorithm lead to a payment of at least €0.60 in 70% of cases.

Table A-2: Distribution of the random influence, which has an effect on the share price

Level of the random influence	Probability
-€30 ≤ Ɛ ≤ -€21 and €21 ≤ Ɛ ≤ €30	5% each (10%)
-€20 ≤ Ɛ ≤ -€16 and €16 ≤ Ɛ ≤ €20	10% each (20%)
-€15 ≤ Ɛ ≤ -€11 and €11 ≤ Ɛ ≤ €15	20% each (40%)
-€10 ≤ Ɛ ≤ -€6 and €6 ≤ Ɛ ≤ €10	10% each (20%)
-€5 ≤ Ɛ < €0 and €0 ≤ Ɛ ≤ €5	5% each (10%)

As the level of the random influence is not known when handing in a forecast, the optimal strategy is to insert the values of the influencing factors A, B, C and D into the formula for the price formation mechanism and to assume an expected value of 0 for the random influence. This is precisely what the algorithm does. With the information available, it is thus not possible to make better forecasts than the algorithm.

When they make their own forecasts, the subjects also have the additional disadvantage that they do not know the exact formula for the price formation mechanism. They can only create an approximate picture of the price formation mechanism on the basis of examples of rounds of the game for which no payments were made (price history). For this purpose they are provided with the exact level of the share price, the influencing factors A, B, C and D as well as the random influence from ten previous rounds. From this information it is also already clear that making naïve forecasts – i.e., using the current price K_t without adaptation as a forecast for the following period P_{t+1} – and continuously forecasting the average price of the last ten rounds are not promising approaches.

Given the advantage which the algorithm has in terms of information, there is thus no reason to presume that the subjects could succeed in making better forecasts. In effect they achieve an average total payment of €8.94 with their approach. They are thus clearly behind the payment of €10.03 obtained with the algorithm (p-value Wilcoxon rank-sum test ≤ 0.001). Decisions against using the algorithm can thus be considered to be algorithm aversion.

Chapter V

Impact of the Decoy Effect on Algorithm Aversion

Sole Author
Contribution Marco Lorenz: 100%

Submitted to:
Management Science - Special Issue on The Human-Algorithm Connection

Published in:
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 22-02, Wolfsburg, July 2021. (Working Paper)

Abstract

Limitations in the human decision-making process restrict the technological potential of algorithms, which is also referred to as "algorithm aversion." This study uses a laboratory experiment with test subjects to investigate whether a phenomenon known since 1982 as the "decoy effect" is suitable for reducing algorithm aversion. For numerous analog products, such as cars, drinks, or newspaper subscriptions, the decoy effect is known to have an immense influence on human decision-making behavior. Surprisingly, the decisions between forecasts by humans and robo-advisors (algorithms) investigated in this study are not affected by the decoy effect at all. This is true both a priori and after observing forecast errors.

Keywords

Algorithm aversion, decoy effect, robo-advisors, technology adoption, human-computer interaction, experiments, behavioral economics.

JEL Classification

D81, D83, D91, G41, O30.

Funding

This research project was funded by the grant „Reinhard Selten-Stipendium“ of the German non-profit association "Gesellschaft für experimentelle Wirtschaftsforschung e.V.” The funding was used exclusively to remunerate the subjects in the laboratory experiment.

1. Introduction

Considerable progress in the field of artificial intelligence (AI) is currently paving the way for numerous promising business models. However, many people have reservations about the automated processes they rely on. These reservations are also referred to as "algorithm aversion" (Mahmud et al., 2022; Burton, Stein & Jensen, 2020; Jussupow, Benbasat & Heinzl, 2020).

Algorithms or AI that work with stochastic processes cannot make exclusively accurate predictions, e.g., about the development of capital markets. As soon as users realize this, they often distrust the technology and refrain from using it (Dietvorst, Simmons & Massey, 2015). This problem also occurs when users must decide whether to trust an algorithm or their own judgment. Many users tend to trust themselves even when there is clear evidence that they are unlikely to make better predictions than an algorithm in the long run (Burton, Stein & Jensen, 2020).

Algorithm aversion means that promising technologies do not succeed in the market as one would expect, given their performance and cost advantages. In finance, for example, many users find it difficult to develop trust in automated asset managers, so-called "robo-advisors," even though their use can help avoid costly mistakes (Back, Morana & Spann, 2021). To mitigate this issue, measures to reduce algorithm aversion need to be identified and taken.

An experimental study focusing on this topic investigates what happens when users are given the possibility to manipulate an algorithm’s output. Dietvorst, Simmons & Massey (2018) give some subjects the ability to adjust an algorithm's predictions downward or upward by a few percent in post-processing. This ability to influence an algorithm significantly increases willingness to use. Interestingly, the effect occurs even when the opportunities for adjustment are kept low. The article suggests that algorithm aversion can be mitigated by giving people opportunities to influence the algorithm's predictions, even if only to a limited extent (Dietvorst, Simmons & Massey, 2018).

Other studies have identified shortening response times (Efendić, Van de Calseyde & Evans, 2020) or making the algorithm take into account the predictions of knowledgeable humans (Kawaguchi, 2021), among others, as measures by which algorithm aversion can be reduced. In addition, a more precise representation of the algorithmic output, e.g., by adding additional decimal places (Kim, Giroux & Lee, 2021) or providing information about the procedure and accuracy of an algorithm (Ben David, Resheff & Tron, 2021) also seem to be suitable for this purpose.

However, the listed measures reduce algorithm aversion only to a limited extent. Moreover, many of them include the risk that the prediction quality of the algorithm decreases after human influence (cf. Kawaguchi, 2021; Dietvorst, Simmons & Massey, 2018). Consequently, it remains an important research task to uncover effective ways to reduce algorithm aversion that do not involve degradation of the algorithm's prediction quality.

Therefore, this study takes up an idea of Huber, Payne & Puto (1982), who studied the human decision-making process under uncertainty using economic experiments. Their study shows that the possibility of comparing multiple options can massively influence the human decision-making process. Under certain circumstances, an option finds significantly greater appeal if a comparable

but recognizably inferior option is added. This phenomenon is also known as the "decoy effect" (see chapter 2.1). In numerous replications, the decoy effect has been shown to cause subjects to change their decision behavior when choosing between different consumer goods (Ariely & Wallsten, 1995), services (Park & Kim, 2005), and even lotteries (Kroll & Vogt, 2012; Herne, 1999). These findings suggest that the decoy effect may have an impact on other decision-making situations as well.

In previous studies on algorithm aversion, decision makers were usually provided with only one algorithm, which they could either rely on or refrain from using (e.g., Dietvorst, Simmons & Massey, 2015). As technology advances, the transferability of this experimental design to practice decreases. In fact, in practice we can already choose between several algorithms that differ in their performance for many tasks, not least asset management.

The possibility to choose between several algorithms could lead to a decoy effect whenever one algorithm is obviously better than another in at least one criterion and not worse than that same algorithm in any other criterion. In this case, the objectively superior algorithm would gain additional attractiveness from the user's point of view, even compared to alternative methods in which no algorithm is used.

The great advantage of this approach over measures already identified for reducing algorithm aversion is that the superior algorithm would not have to be manipulated at all. This means that the willingness of decision makers to use the algorithm could be increased without having to compromise the performance or user-friendliness of the algorithm. Therefore, in this study, an economic experiment will be conducted to investigate whether the decoy effect can be used to reduce algorithm aversion. The influence of the decoy effect on the willingness to use an algorithm is investigated both from the outset and after observing errors in the algorithm.

2. Literature overview

2.1 Decoy effect

The decoy effect (or asymmetric dominance effect) was first discovered about 40 years ago. In an economic experiment, Huber, Payne & Puto (1982) let 150 students choose between different cars, restaurants, types of beer, lotteries, movies, and TVs. They found that the addition of a so-called *decoy*, which is comparable to and clearly inferior to the offer referred to as the *target*, can lead to the target being selected significantly more often than a third option, the so-called *competitor*.

The decoy effect is illustrated by Ariely (2009) using the example of subscription offers of the magazine "The Economist." The author divides 200 students into two groups (conditions) and asks them each to estimate which newspaper subscription they would choose if they had to decide between several offers. In the first condition (control), students can choose between the "Digital" ($59) and "Print + Digital" ($125) offers. It turns out that 68 students prefer the "Digital" offer and the remaining 32 study participants prefer the "Print + Digital" offer. If these 100

students actually took out the subscriptions, this would result in total revenue of $8,012 for The Economist (see Table 1).

Table 1: Distribution in the control group in Ariely (2009)

Offer	Digital	Print + Digital	
Price	$59	$125	
Number of buyers	68	32	
Volume of sales per offer	$4,012	$4,000	
Total volume of sales			$8,012

In the second condition (decoy), the offer "Print" ($125) is added to the already known offers "Digital" ($59) and "Print + Digital" ($125). Although the "Print" offer is not selected even once, it has a massive influence on the theoretically achieved sales of "The Economist." In the second group, only 16 students choose the "Digital" offer, while 84 students prefer the "Print + Digital" offer. The total revenue would now be $11,444 (see Table 2).

Table 2: Distribution under the influence of the decoy effect in Ariely (2009)

Offer	Digital	Print + Digital	Print	
Price	$59	$125	$125	
Number of buyers	16	84	0	
Volume of sales per offer	$944	$10,500	$0	
Total volume of sales				$11,444

The "Print" offer is added in the second condition to steer decision-making behavior towards the "Print + Digital" offer. In this context, it is therefore referred to as the *decoy*, while the "Print + Digital" offer, which gains in popularity after the decoy is added, is called the *target*. The "Digital" offer is in turn the *competitor* of "Print." In this context, the target is also said to *asymmetrically dominate* the decoy.

Ariely attributes the massive differences in choice behavior between the "Digital" and "Print + Digital" offers to the addition of the decoy "Print." In the first condition, students had to compare two options in which each choice offered a distinct advantage. The name of the "Print + Digital" option already suggests that this option offers additional consumption possibilities. However, it is obviously inferior to the "Digital" option in terms of price. The individual has to infer whether the added value in the "consumption possibilities" dimension of the "Print + Digital" option should be rated higher than the added value in the "price" dimension of the "Digital" option. Accordingly, the results of consumer preferences vary.

In the second condition, the comparability of the "Print + Digital" and "Print" offers leads the subjects to apply a heuristic. Although here, too, all decisions are exclusively in favor of the

"Digital" and "Print + Digital" options, subjects replace the question of which of the two options is advantageous with a comparison of the "Print + Digital" and "Print" options. At the same price, the target "Print + Digital" clearly offers more consumption possibilities than the decoy "Print." Since the target is clearly superior to the decoy, its appeal is seemingly enhanced, and this also compared to the competitor "Digital," whose relative advantage over other options still cannot be determined.

Park & Kim (2005) investigate whether the decoy effect works in the same way when two decoys are offered. In this case, the first decoy is only asymmetrically dominated by the target. It cannot be easily compared with the competitor. The second decoy, on the other hand, is asymmetrically dominated by both the target and the competitor. It turns out that, with two decoys involved, the target only gains in attractiveness if the participants are asked to first evaluate each of the four options for themselves. If they are asked to compare the options immediately, the decoy effect does not work as usual any longer.

Frederick, Lee & Baskin (2014) raise the question of which measures are most effective in eliciting a decoy effect. They compare the effectiveness of three ways of presenting product dimensions: representation in numerical form, representation in pictorial form, and physical experience of the differences by the subjects themselves (e.g., through the sense of taste). They find that the decoy effect occurs only when the product dimensions are represented as written-down numbers, for example in the form of numerical ratings. Yang & Lynn (2014) also conclude that qualitative-verbal descriptions as well as pictorial representations are not particularly well suited to create an asymmetric dominance effect. The two measures studied lead to significant decoy effects in only 11 of 91 comparisons across 23 different product classes.

Crosetto & Gaudeul (2016) investigate the robustness of the decoy effect. For this purpose, they introduce the so-called "monetary indicator" as an additional dimension. The authors design the monetary indicator in such a way that selecting the target entails higher costs than selecting the competitor. They vary how much higher the cost of selecting the target is compared to the competitor. It turns out that the preference for the target remains as long as it is up to 8% more expensive than the competitor.

Last but not least, the decoy effect has also been shown to have an impact on individuals' risk preference and social behavior. Kroll & Vogt (2012) let subjects choose between different types of lotteries. It is shown that adding a decoy in all cases leads to an increase in participants' risk taking. Wang et al. (2018) have their subjects play a modified form of the "Prisoner’s Dilemma" in which the third option "reward" is added as a decoy to the well-known options "cooperate" and "defect." It can be observed that the decoy leads to subjects cooperating significantly more often and thus also increases their average payoff compared to a control condition.

2.2 Algorithm aversion

Differences in the performance of algorithms and humans in dealing with similar tasks have been studied since the 1950s (Meehl, 1955). In recent years, the interaction between humans and algorithms has been examined more closely, as automated algorithmic activity increasingly

shapes online information and economic systems. In an experimental study, Önkal et al. (2009) find that the same recommendations are followed to a lesser extent when subjects are led to believe that they come from an algorithm than when they are told that the recommendations come from a competent human.

Six years later, Dietvorst, Simmons & Massey (2015) find that errors by an algorithm lead to a greater loss of confidence than errors by a human. Their subjects can delegate a prediction task to either an algorithm or a human. The subjects who were able to watch the algorithm commit errors in trial rounds showed an increasingly dismissive attitude towards the algorithm in the subsequent compensated rounds. To describe this phenomenon, the term "algorithm aversion" was coined. Nowadays, the term is used both to generalize humans' rejection of algorithms and to describe the dramatic diminishment of human trust in algorithmic efficiency upon observing the errors of an algorithm (Filiz et al., 2021a).

Further studies have shown that algorithm aversion is equally observable in different disciplines, such as law (Ireland, 2020), asset management (Niszczota & Kaszás, 2020), medicine (Lennartz et al., 2021), or poetry (Köbis & Mossink, 2021). Its extent nevertheless seems to depend on the context of the task settings (Filiz et al., 2021a; Castelo, Bos & Lehmann, 2019).

Algorithm aversion can occur regardless of whether the users can execute a task themselves or delegate the task to another expert or layperson as opposed to entrusting the work solely to an algorithm (Germann & Merkle, 2020). Unrealistic expectations about the accuracy of an algorithm have been identified as one of the main causes of algorithm aversion (Rebitschek, Gigerenzer & Wagner, 2021). Respondents often assume the error rates of algorithms to be in ranges that are so low that they cannot be achieved in practice. This also partly explains why trust in algorithms drops so rapidly after erroneous predictions (Dietvorst, Simmons & Massey, 2015).

Furthermore, it has been shown that decision makers have reservations about algorithms because, unlike humans, they do not trust them to learn over time and gradually improve their prognosis quality (Berger et al., 2021). Consistent with this is the finding that the so-called "uniqueness neglect" can also be a driver of algorithm aversion. Uniqueness neglect describes the phenomenon in which humans believe each meaningful decision is accompanied by unique circumstances and the intrinsic nature of the decision and surrounding context is unable to be grasped fully by an algorithm (Pezzo & Beckstead, 2020; Longoni, Bonezzi & Morewedge, 2019). Furthermore, humans tend to form stronger emotional bonds with fellow humans than with algorithms, which also leads them to feel hesitant in solely trusting algorithms during collaborative sensemaking processes (Leyer & Schneider, 2019).

Yeomans et al. (2019) also attribute algorithm aversion in part to an overestimation of people's own predictive abilities (overconfidence). By overestimating the likelihood of success when performing a task themselves, people also misjudge the value added by using an algorithm. It has been shown that learning effects, induced by repetitive tasks as well as clear feedback about one's own performance and that of an algorithm, can help to slightly reduce the aversion (Filiz et al., 2021b). Fittingly, algorithm aversion has been observed to decrease with increasing digital literacy of decision makers (Wang, Harper & Zhu, 2020), but to increase with increasing expertise

of decision makers in the field of the particular prediction task (Allen & Choudhury, 2022; Gaube et al., 2021).

In practice, algorithm aversion means that promising innovations based on the use of algorithms and AI do not establish themselves on the market as one would expect in view of their advantages. If algorithm aversion can be overcome, extensive cost savings or the development of new digital business models can be made possible. In recent years, therefore, research has increasingly been devoted to finding measures that are likely to increase decision-makers' confidence in technology and thus their willingness to use it (Ben David, Resheff & Tron, 2021; Kawaguchi, 2021; Kim, Giroux & Lee, 2021; Efendić, Van de Calseyde & Evans, 2020; Dietvorst, Simmons & Massey, 2018). However, the identified measures only reduce algorithm aversion to a small extent and also introduce new problems, such as a deterioration in prediction quality (Kawaguchi, 2021; Dietvorst, Simmons & Massey, 2018).

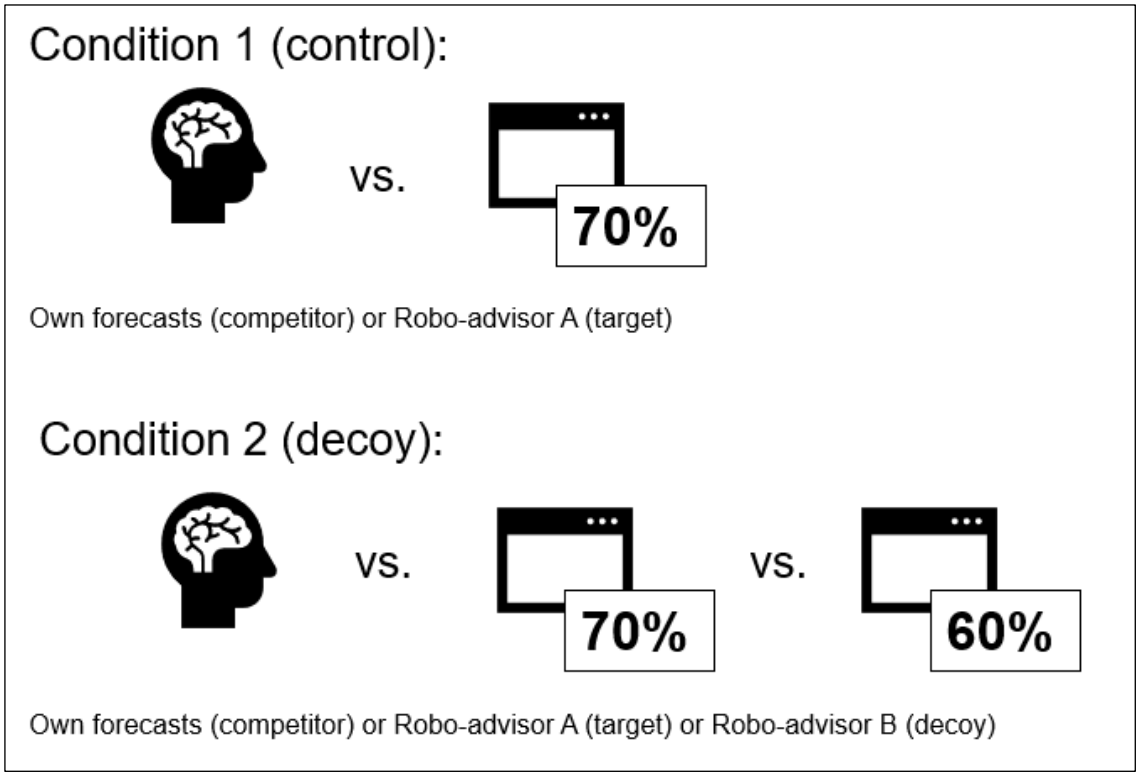
2.3 Hypotheses

The example of an investment experiment is used in order to investigate whether a decoy effect can influence the decision for or against the use of an algorithm. The experimental design of Ariely (2009) described in chapter 2.1 is transferred to the context of an investment decision with automated asset managers, so-called robo-advisors. In condition 1 (control), subjects can perform the investment task independently or delegate it to an algorithm with a success rate of 70%. In condition 2 (decoy), subjects can perform the task independently, delegate to an algorithm with a success rate of 70%, or delegate to a second algorithm with a success rate of 60%. The algorithm with a success rate of 70% will be referred to as "Robo-advisor A" or "Algorithm A" and the algorithm with a success rate of 60% will be referred to as "Robo-advisor B" or "Algorithm B."

By adding the second algorithm, a decoy effect is produced (Figure 1). In this study, Algorithm A acts as the *target*. Algorithm B represents the *decoy* that could increase the attractiveness of the target (Algorithm A). The independent accomplishment of the prediction task by the subjects themselves is the *competitor* in this case. The experiment lasts ten rounds (see chapter 3.2).

Prior research has shown that subjects tend to prefer human judgment over the advice of an algorithm (cf. e.g., Alemanni et al., 2020; Promberger & Baron, 2006). When deciding between algorithms and one's own judgment, this effect can be additionally amplified by an overestimation of one's own forecasting abilities, which is also referred to as "overconfidence" (cf. Filiz et al., 2021b). This is interesting insofar as the probability of achieving a result by human judgment that even comes close to that of a specialized algorithm is extremely low. In view of the previous research results, it can nevertheless be expected that in this study, too, a large proportion of the subjects will rely on their own performance of the task, irrespective of the advantages of the algorithm.

Figure 1: General design of the experiment



Hypothesis 1 is: Not all subjects will choose to always use an algorithm. Null Hypothesis 1 is thus: All subjects will choose to use an algorithm in all of the rounds.

The participants in condition 1 (control) must decide between the better-performing Robo-advisor A with a success rate of 70% (*target*) and the independent asset management by themselves (*competitor*). In condition 2 (decoy), Robo-advisor B with a success rate of 60% is added as a *decoy*, which is comparable to Robo-advisor A and clearly inferior to it.

From a purely mathematical point of view, the decision situation is identical in both cases. It only depends on whether the chances of success are estimated to be higher when performing the prediction task independently than when choosing Robo-advisor A. Robo-advisor B should not have any influence on the decisions of a subject acting as a strictly rational utility maximizer (*Homo Oeconomicus*) because of the lower probability of success.

If the decoy effect sets in analogous to the form observed by Ariely (2009), the comparability of Robo-advisor A with Robo-advisor B with respect to the dimension "success rate" in the second condition will, however, lead to a significantly higher number of decisions in favor of delegation to the superior algorithm and against independent performance of the prediction task.

Hypothesis 2 is: The proportion of decisions in favor of the target algorithm will be higher if another algorithm is introduced as a decoy. Null Hypothesis 2 is: The proportion of decisions in favor of the target algorithm will not be higher if another algorithm is introduced as a decoy.

A crucial aspect of algorithm aversion lies in the reaction to flawed predictions. As in this study, the subjects of Dietvorst, Simmons & Massey (2015) are faced with the choice whether to perform a prediction task independently or to delegate it to an algorithm. Some of the subjects have the opportunity to observe the algorithm in advance as it performs its task (and consequently, inevitably, as it commits errors). The authors investigate to what extent this influences the decision behavior of the subjects. They find that subjects who were able to observe the algorithm making inaccurate predictions actually rely significantly more often on their own judgment than on the algorithm in subsequent rounds, even though the algorithm still has the higher success rate. Interestingly, the effect does not occur to the same extent after observing unsuccessful predictions by a human. This finding is confirmed by Bogert, Schecter & Watson (2021), who also conclude that subjects are more sensitive to errors made by an algorithm than to errors made by a human.

In this study, it will be investigated whether algorithm aversion following erroneous forecasts also occurs when multiple algorithms are available. Adding another algorithm in condition 2 (decoy) could lead to a weakening of the effect, since the subjects now have an additional alternative to choose from. Even if they no longer have sufficient confidence in the algorithm that made an inaccurate prediction, they do not necessarily have to abandon the use of algorithms altogether but can simply use the second algorithm. In addition, the resulting distrust in the algorithm could be weakened by the additional information that it is nevertheless a high-performing algorithm relative to the other algorithm. This would mean that the rejection of algorithms as decision-making aids after an incorrect forecast is no longer so pronounced if several algorithms are available.

Hypothesis 3 is therefore: The proportion of subjects who subsequently use an algorithm after an incorrect prognosis of an algorithm will be significantly higher in condition 2 (decoy) than in condition 1 (control). Null Hypothesis 3 is: The proportion of subjects who subsequently use an algorithm after an incorrect prognosis of an algorithm will not be significantly higher in condition 2 (decoy) than in condition 1 (control).

Based on this, the reaction to one's own inappropriate predictions could also vary if several algorithms are available as alternatives. In contrast to the study by Dietvorst, Simmons & Massey (2015), a broader range of options and additional information on relative performance could increase the willingness to abandon the initial rejection attitude of algorithms after experiencing the difficulty of a prognosis task firsthand.

Hypothesis 4 is: The proportion of subjects who subsequently switch to an algorithm after making an incorrect prognosis themselves will be significantly higher in condition 2 (decoy) than in condition 1 (control). Null Hypothesis 4 is: The proportion of subjects who subsequently switch to an algorithm after making an incorrect prognosis themselves will not be significantly higher in condition 2 (decoy) than in condition 1 (control).

3. Experimental design

3.1 Participants

To answer the research questions, an economic experiment is conducted with students of the Ostfalia University of Applied Sciences in Wolfsburg, Germany. A total of 160 students are taking part in the experiment in the university's research laboratory between April 20, 2022 and April 28, 2022. Each of the total of 28 sessions lasts approximately half an hour. The subjects appear concentrated during the experiment. They receive an average payment of €5.95, which seems to have created an effective incentive for meaningful decision-making.

The subjects are on average 23.6 years old and in the fifth semester of their studies. For 123 participants (78%), it is their first time partaking in an economic experiment, and the remaining 37 participants (22%) have already engaged in other economic experimental studies. The subjects are equally divided into two groups: a treatment condition, in which a decoy is introduced, and a control condition. To avoid bias of the results, an equal distribution with respect to faculty and gender is taken into account (Figures 2 and 3).

Figure 2: Distribution of the subjects' faculties in the conditions

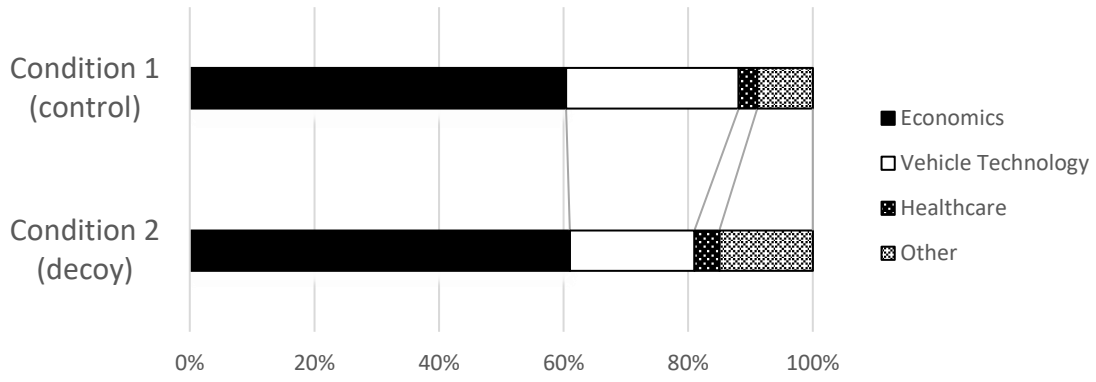
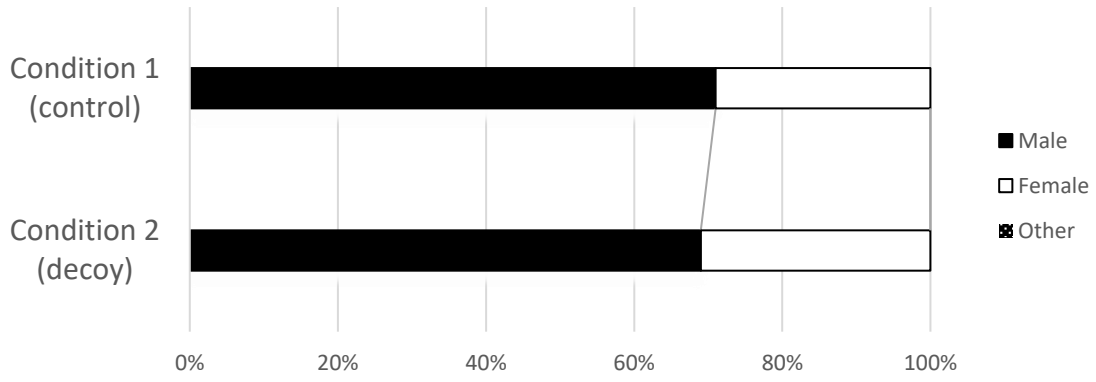


Figure 3: Distribution of the subjects' genders in the conditions



3.2 Task

The subjects have to cope with a task from the field of asset management. In the course of ten game rounds, each subject must decide whether to invest an allocated initial budget of 10 experimental currency units (ECU) in the so-called "Z Share" or to save the budget. In each game round, subjects have to invest or save the full amount of 10 ECU. It is not possible to split the budget within a game round. At the end of the game, the accumulated profit in ECU from the individual game rounds is converted and paid out to the participants as compensation.

The participants' goal is therefore to maximize their credit. To do this, in each of the ten rounds of the game, they can either manage their budget themselves or entrust it to a robo-advisor (control condition) or one of two robo-advisors (decoy condition), who invest or save the budget in their place. The detailed instructions are in Appendix A.

When the budget is invested in the share, the round profit may be above or below the initial budget of 10 ECU due to price fluctuations. For example, if the price of the share increases by 10% within a round, and 10 ECU have been invested in the share, 11 ECU will be credited to the balance. If the share price falls by 10% in one round and 10 ECU are invested in the share, the balance is credited with 9 ECU. If, on the other hand, the decision is made to save, the subjects' balance is always credited with exactly 10 ECU. Before making the investment decision, the subjects should therefore get an idea of whether they think the share price will rise or fall in the respective round. If the price rises, it is the profit-maximizing strategy to invest in the share. If the price falls, the profit-maximizing strategy is to save.

As an aid, before each game round, the current values of four fundamental influencing factors are announced. In combination with a fifth influencing factor (the so-called "random influence"), the four fundamental influencing factors determine the share price trend. The instructions describe in detail in which ranges the values of the influencing factors can vary. In addition, the subjects receive insight into the distribution of the values of the fundamental influencing factors and the random influence in ten rounds of share price history. Based on this information, they can get an approximate picture of the correlations and the current development. The values of the influencing factors in each round were generated once as a random process and are identical for each subject.

The participants in condition 1 (control) are given the choice of performing the asset management task independently or delegating it to Robo-advisor A with a success rate of 70%. That is, the algorithm makes the decision that maximizes the subjects' assets (invest or save) in an average of 7 out of 10 rounds. In the second condition (decoy), in addition to the independent execution and the Robo-advisor A, the subjects also have the Robo-advisor B with a success rate of 60%.

The success probabilities of the algorithms are presented in a table in the instructions, along with other product dimensions. This follows the recommendation of Frederick, Lee & Baskin (2014), who found that the presentation of product dimensions in number form is particularly well suited to elicit a decoy effect.

3.3 Procedure

The experiment is implemented in the experimental software z-Tree (cf. Fischbacher, 2007) and is conducted in the laboratory of the Ostfalia University of Applied Sciences. Subjects participate in the experiment from a computer workstation and receive instructions in paper form. The experiment is moderated throughout by a game leader. This ensures that participants actually take the necessary time, do not use any unauthorized aids, and are not disturbed while they have to focus on the task.

The participants first read the instructions for their respective condition. Subsequently, control questions appear on the screen to check whether they have understood the task and correctly recorded all relevant information (see Appendix B). In the second condition, this also ensures that a decoy effect can occur.

The experiment starts with the subjects being given an insight into ten rounds of share price history as well as the values of the four fundamental influencing factors for the current round for orientation purposes (see Appendix C). At this point, the subjects decide for the first time whether they want to perform the investment task in the current round independently or delegate it to Robo-advisor A (condition 1) or either Robo-advisor A or Robo-advisor B (condition 2). If the subjects decide to complete the task independently, they must also decide whether to invest their budget in the Z Share in the current round (expectation: share price rises) or save it (expectation: share price falls). This concludes the first round of the game.

At the beginning of each new game round, the subjects are given an insight into the development of the share price and the influencing factors in the past ten game rounds, as well as the current values of the fundamental influencing factors. In addition, the change in their cumulative balance in the previous game round is always displayed. This allows the subjects to recognize whether the optimal investment decision was made in the past game round or whether there was a forecast error. In each game round, the subjects can decide again whether they want to perform the investment task independently or use the algorithm (condition 1) or one of the two algorithms (condition 2).

After completion of the tenth game round, the subjects are informed of their total compensation. Next, they answer a short questionnaire asking for demographic information. Subsequently, the payout takes place. From the accumulated balance at the end of the tenth game round, 95 ECU are deducted. The remaining ECU are exchanged in the ratio of 1 ECU = €1 and paid out to the participants as compensation. The higher the accumulated credit from the ten rounds of play, the higher the payout for the subjects.

3.4 The algorithms

The two robo-advisors use the values of the fundamental influencing factors to make a prognosis of how the share price will develop. If their model predicts a rising price, they invest the subjects' budget in the share; otherwise, they save it. The formula that the algorithms rely on is designed to make the decision that is favorable to the subjects in the majority of the rounds of the game,

but they also occasionally miss the mark. In order to analyze the subjects' reactions to incorrect predictions, it is crucial that the experiment uses algorithms whose predictions are not always correct.

Robo-advisor A uses exactly the equation behind the share price formation mechanism. It always enters the current values of the fundamental influencing factors into the formula and makes a forecast on this basis. Only the amount of the random influence, which acts as the fifth influencing factor, is not known to Robo-advisor A. It therefore always calculates with the expected value of the random influence (0). The random influence leads to the fact that the forecasts of the algorithm only in 70% of the cases lead to the financially advantageous decision (invest or save). In 30% of the cases, the random influence reverses the direction of the share price development suggested by the four known fundamental influencing factors.

Robo-advisor B uses the same approach. However, this algorithm has no access to the values of fundamental influencing factor B. Therefore, it always calculates with the mean value of the range of fundamental influencing factor B, which is 15 (see instructions in Appendix A). The use of approximate formulas to predict future values is an established procedure when algorithms lack relevant information (cf. Rencher & Schaalje, 2008).

The formula behind the Z Share price development mechanism is:

$$0.8 \times \text{Fundamental Influencing Factor A} + 0.2 \times \text{Fundamental Influencing Factor B} - 0.4 \times \text{Fundamental Influencing Factor C} + 0.04 \times \text{Fundamental Influencing Factor D} + \text{Random influence}$$

The formula used by Algorithm A is:

$$0.8 \times A + 0.2 \times B - 0.4 \times C + 0.04 \times D + 0$$

The formula used by Algorithm B is:

$$0.8 \times A + 0.2 \times 15 - 0.4 \times C + 0.04 \times D + 0$$

The price formation mechanism and the procedure of the algorithms are shown in Table 3. They are illustrated below using the example of game round 5. In game round 5, the value of fundamental influencing factor A is 12, fundamental influencing factor B is 9, fundamental influencing factor C is 7, fundamental influencing factor D is 30, and the random influence is -1 (see Table 3 – section "Influencing Factors"). Thus, the price of the Z Share takes the following value:

$$0.8 \times 12 + 0.2 \times 9 - 0.4 \times 7 + 0.04 \times 30 + (-1) = 8.80 \text{ ECU}$$

Subjects who invest their round budget in this game round will be credited 8.80 ECU. Subjects who save their round budget will always be credited 10.00 ECU. So, in this round it is advisable to save the round budget of 10 ECU instead of investing it in the Z Share. The delta is -1.20 ECU (see Table 3 – section "Yield in ECU").

Algorithm A uses the following equation in this round of the game:

$$0.8 \times 12 + 0.2 \times 9 - 0.4 \times 7 + 0.04 \times 30 + 0 = 9.80\ ECU$$

Thus, its model predicts that no higher round profit can be achieved by investing (+9.80 ECU) than by saving (+10.00 ECU). The predicted delta between investing and saving is -0.20 ECU. If this delta is negative or exactly 0, i.e., a share price of ≤ 10 ECU is predicted, the algorithm saves the subjects' round budget. Their balance is therefore credited with 10 ECU when choosing Algorithm A in game round 5.

Algorithm B uses the following equation in this round of the game:

$$0.8 \times 12 + 0.2 \times 15 - 0.4 \times 7 + 0.04 \times 30 + 0 = 11.00\ ECU$$

Thus, its model predicts that a higher round profit can be achieved by investing (+11.00 ECU) than by saving (+10.00 ECU). The predicted delta between investing and saving is +1.00 ECU. If this delta is positive, i.e., a share price of > 10 ECU is predicted, the algorithm invests the subjects' round budget in the Z Share. However, the actual share price at the end of the round is only 8.80 ECU. The credit balance of the subjects who bet on Algorithm B in game round 5 is consequently credited with 8.80 ECU (see Table 3 – section "Algorithm Success").

Table 3: Price formation mechanism, forecasts of algorithms, and remuneration depending on chosen strategy

Game Round	1	2	3	4	5	6	7	8	9	10	Sum	Remuneration*
Influencing Factors												
Influencing Factor A	13	11	6	8	12	11	6	11	10	15	-	-
Influencing Factor B	15	13	17	18	9	14	20	18	16	11	-	-
Influencing Factor C	6	8	2	1	7	6	5	8	5	9	-	-
Influencing Factor D	19	27	35	32	30	23	24	22	23	21	-	-
Random Influence	0	1	0	-2	-1	0	2	1	-1	0	-	-
Yield in ECU												
Invest	11.76 ECU	10.28 ECU	8.80 ECU	8.88 ECU	8.80 ECU	10.12 ECU	9.76 ECU	11.08 ECU	9.12 ECU	11.44 ECU	100.04 ECU	-
Save	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	100.00 ECU	-
Delta (Invest - Save)	+1.76 ECU	+0.28 ECU	-1.20 ECU	-1.12 ECU	-1.20 ECU	+0.12 ECU	-0.24 ECU	+1.08 ECU	-0.88 ECU	+1.44 ECU	-	-
Optimal Strategy	Invest	Invest	Save	Save	Save	Invest	Save	Invest	Save	Invest	-	-
Algorithm Success												
Price Forecast (Alg. A)	11.76 ECU	9.28 ECU	8.80 ECU	10.88 ECU	9.80 ECU	10.12 ECU	7.76 ECU	10.08 ECU	10.12 ECU	11.44 ECU	-	-
Price Forecast (Alg. B)	11.76 ECU	9.68 ECU	8.40 ECU	10.28 ECU	11.00 ECU	10.32 ECU	6.76 ECU	9.48 ECU	9.92 ECU	12.24 ECU	-	-
Forecast of Delta (Alg. A)	+1.76 ECU	-0.72 ECU	-1.20 ECU	+0.88 ECU	-0.20 ECU	+0.12 ECU	-2.24 ECU	+0.08 ECU	+0.12 ECU	+1.44 ECU	-	-
Forecast of Delta (Alg. B)	+1.76 ECU	-0.32 ECU	-1.60 ECU	+0.28 ECU	+1.00 ECU	+0.32 ECU	-3.24 ECU	-0.52 ECU	-0.08 ECU	+2.24 ECU	-	-
Decision of Algorithm A	Invest	Save	Save	Invest	Save	Invest	Save	Invest	Invest	Invest	-	-
Decision of Algorithm B	Invest	Save	Save	Invest	Invest	Invest	Save	Save	Save	Invest	-	-
Remuneration												
Only Incorrect Forecasts	10.00 ECU	10.00 ECU	8.80 ECU	8.88 ECU	8.80 ECU	10.00 ECU	9.76 ECU	10.00 ECU	9.12 ECU	10.00 ECU	95.36 ECU	€0.36
Only Correct Forecasts	11.76 ECU	10.28 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.12 ECU	10.00 ECU	11.08 ECU	10.00 ECU	11.44 ECU	104.68 ECU	€9.68
Always Saving	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	10.00 ECU	100.00 ECU	€5.00
Always Investing	11.76 ECU	10.28 ECU	8.80 ECU	8.88 ECU	8.80 ECU	10.12 ECU	9.76 ECU	11.08 ECU	9.12 ECU	11.44 ECU	100.04 ECU	€5.04
Choosing Algorithm A	11.76 ECU	10.00 ECU	10.00 ECU	8.88 ECU	10.00 ECU	10.12 ECU	10.00 ECU	11.08 ECU	9.12 ECU	11.44 ECU	102.40 ECU	€7.40
Choosing Algorithm B	11.76 ECU	10.00 ECU	10.00 ECU	8.88 ECU	8.80 ECU	10.12 ECU	10.00 ECU	10.00 ECU	10.00 ECU	11.44 ECU	101.00 ECU	€6.00

Game rounds in which an algorithm does not make the profit-maximizing decision are highlighted in gray.

*Remuneration is calculated from the cumulative balance at the end of the 10th game round minus 95 ECU, exchanged in the ratio 1 ECU = €1 (see Appendix A).

3.5 Strategies

The subjects have three main strategies at their disposal. They can choose to neglect the robo-advisors and carry out the investment task independently in all ten rounds of the game. Their compensation is then strongly dependent on the success of their forecasts. Based on the values of the fundamental influencing factors generated by a random process, it will range from €0.36 (in the case of ten incorrect forecasts) to €9.68 (in the case of ten correct forecasts).

If subjects save in all ten rounds, their compensation is €5.00 (10 rounds × 10 ECU - 95 ECU). If subjects invest in the Z Share in all ten rounds, they receive €5.04. In three out of ten rounds, the share price trend suggested by the fundamental influencing factors is reversed by the random influence, the amount of which is unknown in advance. Consequently, it cannot be assumed that the subjects will succeed in earning the maximum possible compensation. If subjects always make the investment decision implied by the course of the fundamental influencing factors, their compensation for the ten randomly designed price rounds is €7.40. Thus, the compensation for independent forecasting will presumably be on average in the range between €5.00 (random asset management) and €7.40 (structured asset management based on the fundamental influencing factors).

Furthermore, subjects can use Robo-advisor A in all ten rounds of the game. In this case, their compensation is €7.40 because the algorithm optimally exploits the information content of the fundamental influencing factors.

In the second condition (decoy), the subjects can also use Robo-advisor B throughout. In this case, their compensation is below the compensation when choosing the superior Robo-advisor A and amounts to €6.00 (see Table 3 - section "Remuneration").

While human subjects can only roughly estimate the price formation mechanism on the basis of historical data on price development, the algorithms know the exact price formation mechanism and also have substantial advantages in linking the given information. In order to achieve the success probability of Algorithm A, participants would have to evaluate all information from the ten rounds of price history optimally. To do this, they would have to analyze the effects of each of the four fundamental influencing factors as well as the random influence on the share price and derive the price formation mechanism by an extremely complex regression equation. They would then have to enter the values of the fundamental influencing factors in each of the remunerated game rounds into the formula of the price formation mechanism and, on the basis of the result, make a decision as to whether to save or invest their budget.

But even if they succeeded in doing so, the expected value of their remuneration would only be the same as when using Algorithm A, which also makes optimal use of all the information available in advance. In order to exceed the probability of success of Algorithm A, on top of that, the subjects would have to guess correctly in which rounds of the game the random influence, the amount of which is unknown in advance, would cause a change in the sign of the share price. To beat the algorithm in this experiment, therefore, not only outstanding analytical skills are required, but also a large amount of luck. This is exactly why algorithm aversion has attracted the interest of behavioral economists. Decisions against an algorithm that is so superior, and that are also associated with financial disadvantages, are often placed in the context of cognitive biases.

3.6 Methods

To test hypothesis 1 (Not all subjects will choose to always use an algorithm), the decisions in favor of the target algorithm and decoy algorithm in all ten game rounds are added up for each subject, regardless of the treatment. Then, using the one-sample t-test, it is checked whether the number of game rounds in which an average subject relies on an algorithm is significantly different from 10 out of 10 game rounds (100%). In addition, the Z-test is used to determine whether the proportion of subjects who consistently bet on the algorithm differs significantly from 100% of the subjects (160 out of 160).

To test hypothesis 2 (The proportion of decisions in favor of the target algorithm will be higher if another algorithm is introduced as a decoy), the mean value of the decisions in favor of the target algorithm is determined in both conditions. Using the Wilcoxon rank sum test, it can be checked whether there is a significant difference between the conditions.

Hypothesis 3 is: The proportion of subjects who subsequently use an algorithm after an incorrect prognosis of an algorithm will be significantly higher in condition 2 (decoy) than in condition 1 (control). Here the chi-square test is used. It checks whether the proportion of decisions in favor of the algorithms differs significantly between the conditions. For hypothesis 3, all situations are selected in which a subject delegated the decision to an algorithm in any round between game round 1 and game round 9 and the algorithm did not make the profit-maximizing decision (invest or save), i.e., the algorithm made an error. Subsequently, for both conditions it is separately recorded in how many cases an algorithm was selected again in the following game round and in how many cases the subjects made the investment decision themselves in the following game round. The resulting 2x2 contingency table (condition 1 vs. condition 2; own execution vs. algorithm) is subjected to the chi-square test.

In addition, the same procedure is applied solely to the responses to the first (second, ..., n-th) error of an algorithm that a subject observes. Again, the chi-square test is used to check whether the decisions in the follow-up round differ significantly between the treatments. This additional procedure has the advantage that each subject is included only once in each chi-square test. This can lead to possible biases in the results, e.g., due to differentially pronounced learning effects, being less significant.

Hypothesis 4 is: The proportion of subjects who subsequently switch to an algorithm after making an incorrect prognosis themselves will be significantly higher in condition 2 (decoy) than in condition 1 (control). To test this hypothesis, the same procedure is used as for hypothesis 3, with the only difference that now only the game rounds after incorrect predictions by the subjects themselves (instead of incorrect predictions by the algorithms) are taken into account.

4. Results

4.1 General

160 subjects each make 10 decisions between independent asset management and delegation of the task to an algorithm. In total, 1,600 decisions are observed. Of these, 899 (56.188%) are for

independent asset management and only 701 (43.813%) are for one of the two algorithms. Subjects who manage their assets independently invest their round budget in Z Shares in 577 cases (64.182%) and save the budget in 322 cases (35.818%).

The 43.813% of decisions for the algorithms are divided into 679 decisions (42.438%) in favor of Algorithm A with a success rate of 70% and 22 decisions (1.375%) in favor of Algorithm B with a success rate of 60%. Age, gender, and faculty of the subjects have no significant influence on the decision between algorithm and independent asset management. Moreover, the distribution remains constant over the course of the experiment. In every single one of the ten rounds of the game, only between 35% (round 4) and 49% (round 10) of the subjects decide to use an algorithm, despite its obvious advantages (see Section 3.5). The participants in the experiment are clearly subject to the phenomenon of algorithm aversion. Only 18 of the 160 subjects consistently rely on an algorithm (p-value Z-test < 0.001). The one-sample t-test supports that subjects are far from selecting an algorithm in all ten rounds of the game ($t = -21.376$, $p < 0.001$). The 95% confidence interval ranges from 3.862 to 4.900 out of 10 decisions per algorithm per subject.

In terms of the number of correct predictions, the subjects are clearly inferior to both algorithms. The independent asset management by the subjects leads to the profit maximizing investment decision in 43.604% of the cases (392 out of 899 decisions) and in 56.396% of the cases (507 out of 899 decisions) not to the profit maximizing investment decision. As expected, the complexity of the task poses considerable problems for the subjects in independent asset management. Their success rate is even below 50%.

In the instructions, the probability of success of Algorithm A was given as 70%. Its predictions are correct in 71.429% of the actual observed cases (485 of 679 decisions). Recommendations of Algorithm B are correct in 63.636% of the actually observed cases (14 out of 22 decisions), which also fits the reported success probability of 60%.

Consistent with these results, subjects who consistently make their own predictions achieve a compensation of €5.12 on average. In comparison, the average compensation of subjects who consistently rely on an algorithm is €7.27 (Table 4). Linear regression analysis shows that the compensation increases on average by 19.793 cents with each additional decision in favor of one of the two algorithms (p-value < 0.001).

In view of these figures, the rejection attitude toward the algorithm is remarkable. Over the course of ten rounds of play, subjects pay for their algorithm aversion with a reduction in their average remuneration by up to €1.98, or up to 30% on average. This is nevertheless in line with previous studies, in which algorithm aversion also occurs, although the renunciation of the algorithm drastically reduces the expected value of the compensation.

Table 4: Average remuneration depending on the frequency with which the algorithm was selected

Number of Rounds in Which an Algorithm Was Selected	Number of Subjects	Ø Remuneration
0	26	€5.12
1	15	€5.65
2	16	€5.91
3	17	€5.60
4	11	€5.16
5	15	€5.14
6	12	€6.16
7	15	€6.79
8	8	€6.84
9	7	€7.08
10	18	€7.27
	160	€5.95

4.2 Differences between conditions

Out of 800 decisions in the control condition, 338 (42.250%) are made in favor of the target algorithm and 462 (57.750%) in favor of the independent asset management. In the decoy condition, 341 decisions (42.625%) are made in favor of the target algorithm, 437 decisions (54.625%) are made in favor of the independent asset management, and 22 decisions (2.750%) are made in favor of the decoy algorithm (Figure 4 and Table 5).

Figure 4: Comparison of the conditions

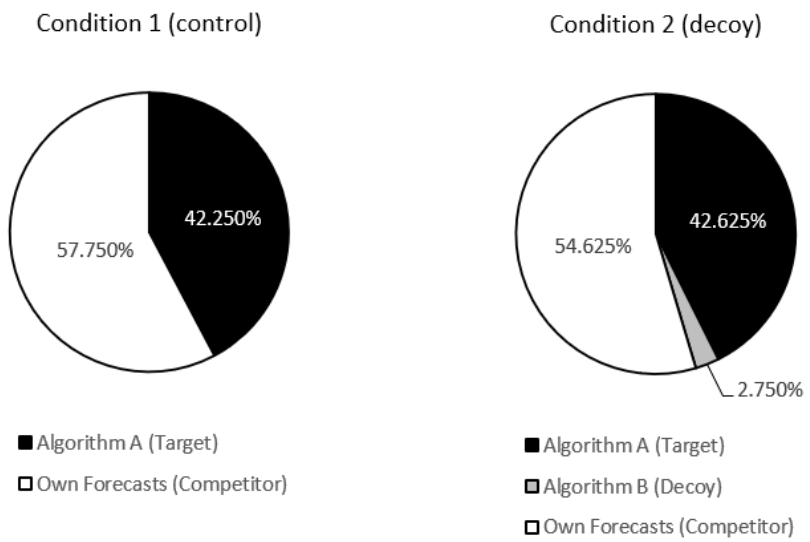


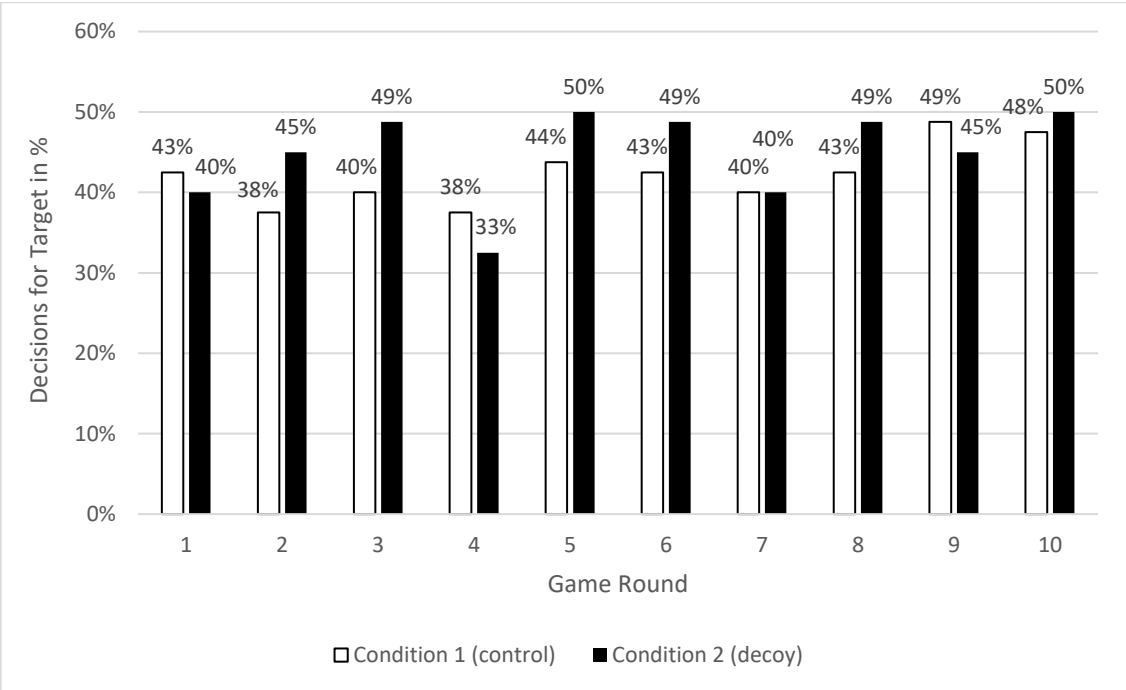
Table 5: Comparison of the conditions

Condition	Own Forecasts	Algorithm A	Algorithm B	Total Algorithm
Control (1)	462 (57.750%)	338 (42.250%)	-	338 (42.250%)
Decoy (2)	437 (54.625%)	341 (42.625%)	22 (2.750%)	363 (45.375%)
Total	899 (56.188%)	679 (42.438%)	22 (1.375%)	701 (43.813%)

The proportion of decisions in favor of own predictions thus decreases slightly in condition 2 (decoy). However, it is by no means the case that subjects now massively select target Algorithm A, as previous research on the decoy effect might have suggested. Of the 25 additional decisions made in favor of the algorithms in the decoy condition, 22 are for Algorithm B (decoy) and only 3 are for Algorithm A (target). The p-value of the Wilcoxon rank sum test of 0.889 proves that the addition of the decoy in the form of Algorithm B in condition 2 does not lead to any significant increase in the use of the target algorithm.

It can be argued just as little that adding a decoy contributes to the reduction of algorithm aversion. Aggregating the decisions in favor of the two algorithms, the difference between the conditions is still far from significant (p-value of the Wilcoxon rank sum test = 0.530). The observed difference of 25 decisions (3.125%) is too small to constitute a significant result. It is rather due to noise in the decisions of the eleven subjects who, despite its lower probability of success, select Algorithm B once or several times.

Figure 5: Comparison of the decisions in favor of the target algorithm between the conditions per game round



Further analyses show that the behavior of the subjects is hardly influenced by the introduction of a decoy. Firstly, the number of subjects who consistently follow a certain strategy is almost identical in both conditions. 18 subjects choose the algorithm in all ten rounds of the game. They are evenly distributed between both conditions (9 each). Of the 26 subjects who do not select the algorithm in any single game round, 12 are in condition 1 (control) and 14 in condition 2 (decoy). Secondly, there are no differences in the time course of the game (Figure 5). The difference in the frequency with which the target algorithm is selected is always in the small range of 0% (round 7) to 8.750% (round 3).

In line with all these findings, the average compensation in the two conditions is also close to each other. The average compensation per subject is €6.02 in condition 1 (control) and €5.89 in condition 2 (decoy). The difference is not significant in the Wilcoxon rank sum test (p-value = 0.371).

4.3 Reaction to forecast errors

Of the 1,600 total forecasts made, 891 are correct and 709 are incorrect. The latter are divided into 507 incorrect forecasts by subjects themselves and 202 incorrect forecasts by one of the algorithms. In the following, the reaction to incorrect predictions in the first nine game rounds is examined, since only these game rounds are followed by at least one more game round in which a change in behavior is possible.

The algorithm makes 202 incorrect predictions in the first nine game rounds: 99 in condition 1 (control) and 103 in condition 2 (decoy). In the control group, 69.697% of subjects continue to use the algorithm in the subsequent round regardless. 30.303% of the subjects withdraw their trust in the algorithm immediately after an error and opt to make their own investment decision in the subsequent round. This result is consistent with previous studies that found that human trust in algorithms declines rapidly after erroneous predictions (cf. Dietvorst, Simmons & Massey, 2015).

How will subjects react if not only another algorithm is available but also the decoy effect provides additional evidence for the relatively good performance of the target algorithm? In condition 2 (decoy), only 62.136% of the subjects choose the target algorithm immediately after they observe an error of an algorithm. 4.854% of the subjects subsequently choose the decoy algorithm and 33.010% choose the independent prediction by themselves (Table 6).

Thus, the expected effect does not occur. In fact, the extent of algorithm aversion after forecast errors of an algorithm even seems to slightly increase under the impression of a decoy effect. The p-value in the chi-square test that includes all decisions made after observing an error of an algorithm is 0.679, which is not a significant result. When only the response to the first (second, ..., n-th error) of an algorithm is considered, the difference is also not significant (first error of an algorithm: n = 105, p = 0.781; second error of an algorithm: n = 66, p = 0.421; third error of an algorithm: n = 29, p = 0.453; for more than three errors of the algorithm, the sample size is less than 20 participants).

Table 6: Responses to forecast errors by an algorithm

Selection in next round	Condition 1 (control) Own Forecasts	Condition 1 (control) Algorithm A	Condition 1 (control) Algorithm B	Condition 2 (decoy) Own Forecasts	Condition 2 (decoy) Algorithm A	Condition 2 (decoy) Algorithm B
Total	30	69	-	34	64	5
Percentage	30.303%	69.697%	-	33.010%	62.136%	4.854%

The subjects' own investment decisions lead to 474 errors in the first nine rounds of the game: 230 in condition 1 (control) and 244 in condition 2 (decoy). In the control group, 69.130% of the subjects maintain their strategy of making their own predictions after an error. In contrast, 30.870% switch to the target algorithm after making their own forecast errors in the following round. In condition 2 (decoy), the values are almost identical (Table 7). Here, 69.262% of the subjects continue to give their own forecasts, 29.098% rely on the target algorithm and 1.639% on the decoy algorithm (total algorithm = 30.737%). This difference is also clearly not significant in the chi-square test that includes all decisions made after observing an error of oneself (p-value = 0.975). When only the response to the first (second, ..., n-th error) of a subject itself is considered, the difference between treatments is still not significant (first error of a subject: n = 140, p = 0.601; second error of a subject: n = 120, p = 0.266; third error of a subject: n = 95, p = 0.596; fourth error of a subject: n = 68, p = 0.195; fifth error of a subject: n = 32, p = 0.414; for more than five errors of a subject, the sample size is less than 20).

Table 7: Responses to forecast errors by the subjects themselves

Selection in next round	Condition 1 (control) Own Forecasts	Condition 1 (control) Algorithm A	Condition 1 (control) Algorithm B	Condition 2 (decoy) Own Forecasts	Condition 2 (decoy) Algorithm A	Condition 2 (decoy) Algorithm B
Total	159	71	-	169	71	4
Percentage	69.130%	30.870%	-	69.262%	29.098%	1.639%

It can thus be stated that the decoy is particularly popular when one of the other two options made an error in the previous round. However, the extent of algorithm aversion after prediction errors remains unaffected by the decoy effect both in the case of errors of the algorithm and in the case of errors of the subjects themselves.

5. Discussion

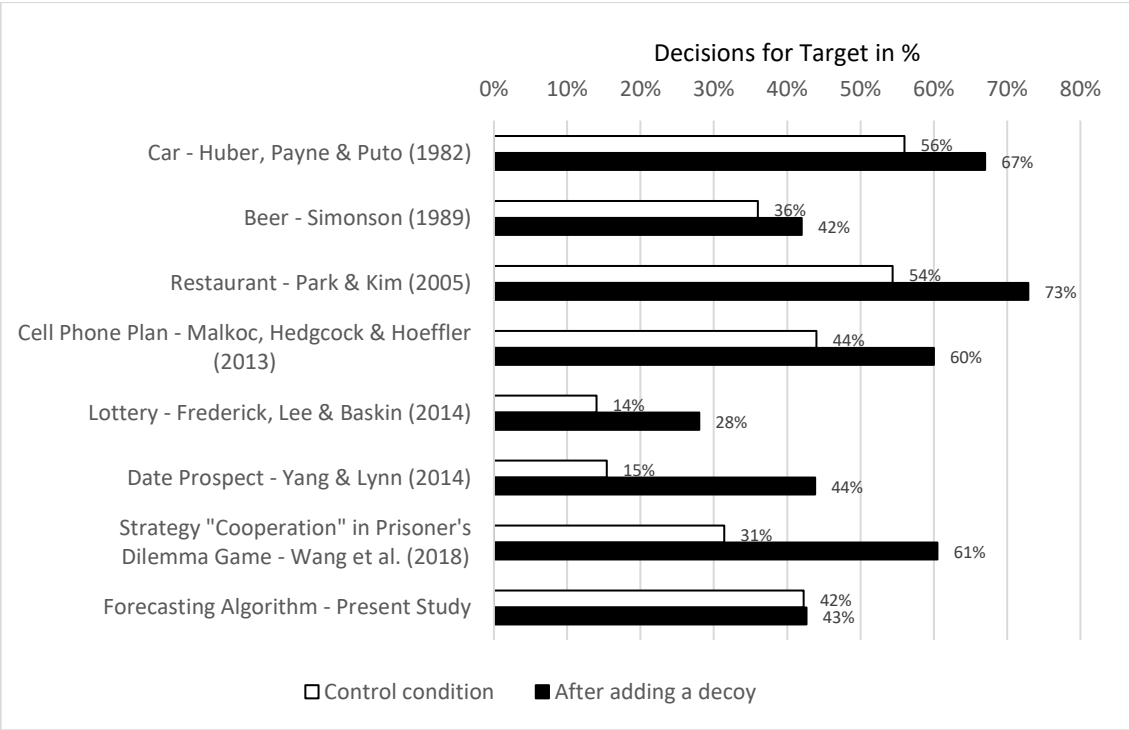
Imagine you are faced with the decision of whether to entrust your private assets to an innovative robo-advisor that promises to increase your wealth. How easy is it to develop trust in the new technology when so much is at stake? The robo-advisor has been explicitly designed to exploit information from the market to maximize your wealth and seems superior to all alternatives. But

will the robo-advisor also succeed in your case, or might you not be better off taking the management of your assets into your own hands?

Now imagine you could compare two robo-advisors available on the market and freely choose between them. It quickly becomes apparent that one of the two robo-advisors seems to be particularly powerful compared to the other. How do you make your initial decision now? Will you succeed in completely ignoring the second robo-advisor in your decision-making process, or will the added option unnecessarily increase the appeal of the superior robo-advisor due to comparison bias?

The results of this study provide initial evidence that the presence of the second robo-advisor makes little difference. The influence of the decoy effect on our decision-making behavior has been demonstrated for numerous products and services from the analog world. The findings so far suggest that adding a decoy should lead to the target gaining greater popularity. Surprisingly, this does not seem to hold true for decoys in the context of novel, complex, digital technologies affected by the phenomenon of algorithm aversion. It has already been shown that emphasizing the statistical superiority of algorithms alone is not sufficient to address algorithm aversion (e.g., Filiz et al., 2021a). The results of this study support the persistence of algorithm aversion. It is apparently so robust that even the otherwise highly reliable decoy effect is unable to mitigate it (Figure 6).

Figure 6: Comparison of decisions in favor of the target product before and after adding a decoy in prior studies



Where multiple experiments with different designs were conducted in the studies, the results from the research design closest to the original design by Huber, Payne & Puto (1982) and the present study are presented here.

In the present study, the decoy effect was made abundantly salient. In the instructions, subjects were given a comparison of the target algorithm with the decoy algorithm in tabular form with easily comparable numbers (Appendix A). Before starting the experiment, all subjects in the decoy condition had to show in the control questions that they understood that two different algorithms were available to them (Appendix B). Last but not least, the success probabilities of the target and the decoy were displayed again in the selection area of the screen in every single round of the game (Appendix C). The fact that the decoy was actually selected in 2.750% of the decisions also fits with the results of previous studies (e.g., Yang & Lynn, 2014; Huber, Payne & Puto, 1982) and suggests that the decoy was noted as a decision option.

However, the presence of the decoy does not lead to an increase in the proportions of the target, as is usually the case. Furthermore, the introduction of the decoy is not suitable for reducing algorithm aversion. Even if the decisions for both algorithms are added together, there is hardly any difference to the control group. This is extremely surprising. On the one hand, the subjects seem to act so rationally that they are hardly influenced by the inferior decoy in their decision between target and competitor. On the other hand, more than half of the decisions are made to perform the prediction task independently, even though the performance in doing so falls significantly short of that of an algorithm.

The main finding of this study is that algorithm aversion prevails over the decoy effect as soon as both have an impact on a decision situation. This is true both throughout the course of the game and explicitly after observing errors of the algorithm. Dietvorst, Simmons & Massey (2015) had uncovered that trust in an algorithm declines rapidly after erroneous predictions, leading to the original coining of the term "algorithm aversion." The results of the present study suggest that this finding is not only valid in the context of one human vs. one algorithm. When an additional decoy algorithm is available, the willingness to use algorithms after errors still declines to the same extent.

In many application areas for algorithms and AI, we are currently at the point where the first offerings are entering the market. For example, autonomous robo-taxis are expected to be offered to the public in Germany for the first time in 2022. For providers, this raises the question of how to raise their chances of success. Previous theory on the decoy effect had implied that providers should offer a decoy in addition to their target product, which they want to establish on the market in the long term, in order to influence potential customers in favor of the target.

However, the results of this study show that the market share of new technologies cannot be increased so easily. Rather, potential users harbor a great deal of skepticism toward innovative, automated processes, which cannot be remedied by adding a decoy to the offering. The pioneers of digitized and automated business ideas should therefore be advised not to pursue the decoy effect as a sales strategy. Instead, they should rather refer to already identified measures to reduce algorithm aversion, such as influencing algorithmic output (Dietvorst, Simmons & Massey, 2018) or learning effects (Filiz et al., 2021b).

Finally, some aspects should be mentioned that may limit the validity of this study for decisions in practice. In order to follow the established research on the decoy effect, the participants in the economic experiment were provided with only two algorithms (target and decoy). In practice we

can often choose between more than two offers that dominate each other in different ways with respect to different dimensions. In particular, the rapid scalability of digital technologies means that the choice usually quickly exceeds two algorithms. Second, it should be mentioned that the results were obtained in the context of robo-advisors. However, asset management is only one small area affected by algorithm aversion. Possibly, different results would be obtained when using other algorithms from areas such as, for example, medicine, transportation, or entertainment. Last but not least, this study did not focus on social influences. However, humans are social beings. In our everyday lives, in contrast to a laboratory experiment, there is a great deal of interaction with other people, which influences our decision-making behavior. It must be left to subsequent research to analyze these aspects in more detail.

6. Summary

In this study, the impact of the decoy effect on algorithm aversion is investigated by means of an economic laboratory experiment. Subjects are divided into two groups for an investment game in which they try to maximize their compensation. In the control condition, they have the choice in ten rounds of the game whether to delegate a prediction task to a specialized algorithm with a success rate of 70% (target) or to handle it independently by themselves (competitor). In the treatment condition, they also have a second algorithm (decoy) available to them in addition to the first algorithm and their own predictions. The second algorithm is identical to the first with one exception: it has a significantly lower success rate of just 60%. We speak of the decoy effect if one option (decoy) is inferior to another option (target) in at least one dimension and superior in no other dimension.

The theory of the decoy effect suggests that the first algorithm (target) should be selected significantly more often in the treatment condition than in the control condition. Once the decoy comes into play, decision makers regularly apply a heuristic. They compare target and decoy and decide in favor of the target, since it is clearly superior to the decoy. In this case, the competitor always loses shares in favor of the target, since adding the decoy does not provide any additional information about comparative effectiveness of the competitor, but only about the target.

In contrast to these considerations is the algorithm aversion. It describes users' reservations about automated procedures (algorithms) that cannot be easily remedied. If users are subject to algorithm aversion, they should not be influenced by the presence of a decoy, because algorithms are generally not an attractive option for them.

The first thing that emerges is that the subjects in this study are also affected by algorithm aversion. Although each decision in favor of one of the two algorithms increases their compensation by an average of 19.793 cents, an algorithm is selected in just 43.813% of the decisions.

Further, the presence of a decoy is shown to have no effect on the extent of algorithm aversion. The proportion of decisions in favor of the better performing target algorithm increases by only 0.375 percentage points after the decoy is added, from 42.250% to 42.625%. Another 2.750% of the decisions are now made in favor of the inferior algorithm (decoy). The proportion of own

predictions by the subjects themselves decreases slightly from 57.750% to 54.625%. The difference turns out to be not statistically significant.

Finally, the reaction to erroneous forecasts is also examined. In slightly more than 30% of the cases, the reaction to forecast errors of the algorithm is to switch to independent forecasting by oneself in the following game round. However, this proportion differs only minimally between conditions. Behavior after algorithm errors is not affected by the decoy effect. The same is true for the behavior after participants' own erroneous forecasts. The proportion of switches to the algorithm is approximately the same in both conditions.

Whether an additional algorithm (decoy) is present or not does not change the willingness to resort to a specialized algorithm in all cases studied. Algorithm aversion cannot be effectively reduced by the decoy effect.

References

Alemanni, B., Angelovski, A., di Cagno, D. T., Galliera, A., Linciano, N., Marazzi, F., & Soccorso, P. (2020). Do Investors Rely on Robots? Evidence from an Experimental Study, *CONSOB Fintech Series, 7,* 1-61.

Allen, R., & Choudhury, P. (2022). Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion, *Organization Science*, *33*(1), 149-169.

Ariely, D., (2009). Predictably Irrational: The Hidden Forces that Shape Our Decisions, HarperCollins, New York City.

Ariely, D., & Wallsten, T. S. (1995). Seeking Subjective Dominance in Multidimensional Space: An Explanation of the Asymmetric Dominance Effect, *Organizational Behavior and Human Decision Processes*, *63*(3), 223-232.

Back, C., Morana, S., & Spann, M. (2021). Do Robo-Advisors Make Us Better Investors?, Discussion Paper No. 276, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 – Rationality and Competition, München und Berlin, http://hdl.handle.net/10419/233499 (accessed on August 10th, 2022).

Ben David, D., Resheff, Y. S., & Tron, T. (2021). Explainable AI and Adoption of Financial Algorithmic Advisors: An Experimental Study, *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,* 390-400.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve - Algorithm Aversion and Demonstrating the Ability to Learn, *Business & Information Systems Engineering*, *63*(1), 55-68.

Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult, *Scientific Reports*, *11*(1), 1-9.

Burton, J., Stein, M. & Jensen, T. (2020). A systematic review of algorithm aversion in augmented decision making, *Journal of Behavioral Decision Making*, *33*(2), 220-239.

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

Crosetto, P., & Gaudeul, A. (2016). A monetary measure of the strength and robustness of the attraction effect, *Economics Letters*, *149*, 38-43.

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, 144*(1), 114-126.

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*(C), 103-114.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021a). The Tragedy of Algorithm Aversion, *Wolfsburg Working Papers 21-02,* Wolfsburg.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021b). Reducing algorithm aversion through experience, *Journal of Behavioral and Experimental Finance, 31*, 100524.

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

Frederick, S., Lee, L., & Baskin, E. (2014). The Limits of Attraction, *Journal of Marketing Research*, *51*(4), 487-507.

Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E. & Ghassemi, M. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids, *npj Digital Medicine*, *4*(1), 1-8.

Germann, M., & Merkle, C. (2020). Algorithm Aversion in Financial Investing, Working Paper, https://dx.doi.org/10.2139/ssrn.3364850 (accessed on August 10th, 2022).

Herne, K. (1999). The Effects of Decoy Gambles on Individual Choice, *Experimental Economics*, *2*(1), 31-40.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis, *Journal of consumer research*, *9*(1), 90-98.

Ireland, L. (2020). Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors, *Journal of Crime and Justice*, *43*(2), 174-192.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion, *Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1-16.

Kawaguchi, K. (2021). When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business, *Management Science*, *67*(3), 1670-1695.

Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations, *Psychology & Marketing*, *38*, 1140-1155.

Köbis, N. & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry, *Computers in Human Behavior*, *114*, 1-13.

Kroll, E. B., & Vogt, B. (2012). The relevance of irrelevant alternatives, *Economics Letters*, *115*(3), 435-437.

Lennartz, S., Dratsch, T., Zopfs, D., Persigehl, T., Maintz, D., Hokamp, N. G., & Dos Santos, D. P. (2021). Use and Control of Artificial Intelligence in Patients Across the Medical Workflow: Single-Center Questionnaire Study of Patient Perspectives, *Journal of Medical Internet Research*, *23*(2), e24221, 1-10.

Leyer, M., & Schneider, S. (2019). Me, you or AI? How do we feel about delegation, *Twenty-Seventh European Conference on Information Systems (ECIS2019),* Stockholm-Uppsala, Sweden, https://aisel.aisnet.org/ecis2019_rp/36 (accessed on August 10th, 2022).

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to Medical Artificial Intelligence, *Journal of Consumer Research*, *46*(4), 629-650.

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion, *Technological Forecasting and Social Change*, *175*, 121390, 1-26.

Malkoc, S. A., Hedgcock, W., & Hoeffler, S. (2013). Between a rock and a hard place: The failure of the attraction effect among unattractive alternatives, *Journal of Consumer Psychology*, *23*(3), 317-329.

Meehl, P. (1955). Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence, University of Minnesota Press, Minneapolis.

Niszczota, P. & Kaszás, D. (2020). Robo-investment aversion, *PLOS ONE*, *15*(9), 1-19.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S. & Pollock, A. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments, *Journal of Behavioral Decision Making*, *22*(4), 390-409.

Park, J., & Kim, J. (2005). The Effects of Decoys on Preference Shifts: The Role of Attractiveness and Providing Justification, *Journal of Consumer Psychology*, *15*(2), 94-107.

Pezzo, M. V., & Beckstead, J. W. (2020). Algorithm aversion is too often presented as though it were non-compensatory: A reply to Longoni et al. (2020), *Judgment and Decision Making*, *15*(3), 449.

Promberger, M., & Baron, J. (2006). Do patients trust computers?, *Journal of Behavioral Decision Making*, *19*(5), 455-468.

Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors by algorithms for credit scoring and recidivism but tolerate even fewer errors, Preprint, https://europepmc.org/article/ppr/ppr321551 (accessed on August 10th, 2022).

Rencher, A. C., & Schaalje, G. B. (2008). Linear Models in Statistics, 2nd Edition, John Wiley & Sons, Hoboken, New Jersey.

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects, *Journal of Consumer Research*, *16*(2), 158-174.

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Paper 684, 1-14.

Wang, Z., Jusup, M., Shi, L., Lee, J. H., Iwasa, Y., & Boccaletti, S. (2018). Exploiting a cognitive bias promotes cooperation in social dilemma experiments, *Nature Communications*, *9*(1), 1-7.

Yang, S., & Lynn, M. (2014). More Evidence Challenging the Robustness and Usefulness of the Attraction Effect, *Journal of Marketing Research*, *51*(4), 508-513.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations, *Journal of Behavioral Decision Making*, *32*(4), 403-414.

Appendix A: Instructions

Appendix A.1: Instructions for Condition 1 (Control)

The Game

In this game, you are asked to make an investment decision in the course of ten game rounds each. You will be given an initial budget of Experimental Currency Units (ECU) of 10 ECU per round. In each of the ten rounds, you can either invest 10 ECU in Z Shares or save them. You always invest or save the full amount; it is not possible to split the budget within a game round.

If you **invest** the 10 ECU in Z Shares, you buy the shares at the beginning of the period and sell the shares again at the end of the period. The selling price will be credited to your balance. It may be higher or lower than the initial amount of 10 ECU you invested, depending on whether the price of the Z Share increased or decreased during the round.

For example, if the price of the Z Share increases by 10% within a round, and you have invested 10 ECU in the Z Share, your balance will be credited with 11 ECU. If the price of the Z Share falls by 10% within a round, and you have invested 10 ECU in the Z Share, your balance will be credited with 9 ECU. **You can therefore invest in the Z Share specifically in the rounds in which you expect the price to rise.**

The share price of the Z Share always results from **four influencing factors** (see Table A-1) plus a **random influence**. The values of the influencing factors are announced to you before each game round.

Table A-1: Factors influencing the formation of the Z Share price[1]

Influencing Factor	Span	Influence	Impact on Share Price
A	5 to 15	Positive	High
B	5 to 25	Positive	Medium
C	0 to 10	Negative	Medium
D	15 to 35	Positive	Low
Random Influence	-2 to +2	Positive	Medium

Influencing factors **A**, **B**, **D,** and the **random influence** have a positive effect on the share price. This means that if these influencing factors are in the upper range of their span (i.e., above the average of the previous periods), the share price tends to rise during the upcoming game round.

Influencing factor **C** has a negative effect on the share price. That is, if this influencing factor lies in the upper range of its span (i.e., above the average of the previous periods), the share price

[1] In the original instructions, this table was referred to as "Table 1."

tends to fall during the coming round. The influencing factors have varying degrees of impact on the share price (Table A-1).

Alternatively, you can **save** the round budget in the amount of 10 ECU. Your balance will then be credited with 10 ECU for the round in question.

Your credit will be built up over the 10 rounds of play and used to calculate your pay at the end of the game. Regardless of your decisions in the previous rounds, you can always invest or save exactly 10 ECU in each new round.

Choice Between Independent Asset Management and an Algorithm

You can also choose in each game round whether you want to manage your round budget independently by yourself or entrust it to a robo-advisor (algorithm).

If you choose the algorithm, it will either invest or save your round budget of 10 ECU in the respective game round in your place. The algorithm will always decide to invest your ECU if its model predicts a rising share price. If its model predicts a falling share price, it will save your budget in that round.

In the past, it has been shown that in 7 out of 10 cases (70%) the algorithm makes the decision (invest or save) that leads to a higher return.

Remuneration

The pay structure is the same whether you manage your budget independently or entrust it to the algorithm. At the end of the game, your total cumulative balance earned in the ten game rounds is considered. 95 of the originally allocated 100 ECU (10 ECU each in 10 game rounds) will be deducted from your balance. The remaining amount will be exchanged for real money at the ratio of 1 ECU = €1 and paid to you as your remuneration.

Procedure

After reading the instructions and answering the control questions, the first remunerated game round (period 11 of 20) starts on your screen.

At the beginning of each game round, you will see the price development of the Z Share, the development of the influencing factors and the development of the random influence for the last ten game rounds (period 1 to 10), in order to get an idea of the development. In addition, you will always be informed of the current values of the four influencing factors for the respective game round. The value of the random influence, on the other hand, is unknown in advance. Afterwards, you make your decision for the respective game round whether you want to manage your round budget independently or entrust it to the robo-advisor (algorithm).

If you decide to do the investment task on your own, next you will choose whether you want to invest 10 ECU in Z Shares or save them in the particular round. If you decide to use the algorithm, it will make the decision between investing and saving in your place.

After submitting the decision, you will be informed about the development of the Z Share price in any case, regardless of whether your budget was invested or saved. So, you will receive the full information in any case. The achieved return from the investment in Z Shares or the saved amount will be credited to your balance.

A total of ten rounds will be played. After the experiment is completed, you will receive your remuneration, which is calculated according to the scheme described under "Remuneration."

Remarks

- Please keep quiet during the experiment!
- Do not look at your neighbor's screen!
- Apart from a pen and a pocket calculator, no other aids (smartphones, smartwatches, etc.) are permitted.
- Only use the white sheet of paper provided for your notes.

Appendix A.2: Instructions for Condition 2 (Decoy)

The Game

In this game, you are asked to make an investment decision in the course of ten game rounds each. You will be given an initial budget of Experimental Currency Units (ECU) of 10 ECU per round. In each of the ten rounds, you can either invest 10 ECU in Z Shares or save them. You always invest or save the full amount; it is not possible to split the budget within a game round.

If you **invest** the 10 ECU in Z Shares, you buy the shares at the beginning of the period and sell the shares again at the end of the period. The selling price will be credited to your balance. It may be higher or lower than the initial amount of 10 ECU you invested, depending on whether the price of the Z Share increased or decreased during the round.

For example, if the price of the Z Share increases by 10% within a round, and you have invested 10 ECU in the Z Share, your balance will be credited with 11 ECU. If the price of the Z Share falls by 10% within a round, and you have invested 10 ECU in the Z Share, your balance will be credited with 9 ECU. **You can therefore invest in the Z Share specifically in the rounds in which you expect the price to rise.**

The share price of the Z Share always results from **four influencing factors** (see Table A-2) plus a **random influence**. The values of the influencing factors are announced to you before each game round.

Table A-2: Factors influencing the formation of the Z Share price[2]

Influencing Factor	Span	Influence	Impact on Share Price
A	5 to 15	Positive	High
B	5 to 25	Positive	Medium
C	0 to 10	Negative	Medium
D	15 to 35	Positive	Low
Random Influence	-2 to +2	Positive	Medium

Influencing factors **A**, **B**, **D,** and the **random influence** have a positive effect on the share price. This means that if these influencing factors are in the upper range of their span (i.e., above the average of the previous periods), the share price tends to rise during the upcoming game round.

Influencing factor **C** has a negative effect on the share price. That is, if this influencing factor lies in the upper range of its span (i.e., above the average of the previous periods), the share price

[2] In the original instructions, this table was referred to as "Table 1."

tends to fall during the coming round. The influencing factors have varying degrees of impact on the share price (Table A-2).

Alternatively, you can **save** the round budget in the amount of 10 ECU. Your balance will then be credited with 10 ECU for the round in question.

Your credit will be built up over the 10 rounds of play and used to calculate your pay at the end of the game. Regardless of your decisions in the previous rounds, you can always invest or save exactly 10 ECU in each new round.

Choice Between Independent Asset Management and an Algorithm

You can also choose in each game round whether you want to manage your round budget independently by yourself or entrust it to one of two algorithms, so-called robo-advisors (Table A-3).

If you choose one of the algorithms, it will either invest or save your round budget of 10 ECU in the respective game round in your place. The algorithms will always decide to invest your ECU if their models predict a rising share price. If their models predict a falling share price, they will save your budget in that round.

In the past, it has been shown that in 7 out of 10 cases (70%) Algorithm A makes the decision (invest or save) that leads to a higher return. Furthermore, it has been shown that Algorithm B makes the advantageous decision in 6 out of 10 cases (60%).

Table A-3: Properties of the algorithms[3]

Property	Algorithm A	Algorithm B
Year of completion	2022	2022
Manufacturer	Ostfalia Analytics	Ostfalia Analytics
Probability of success	70%	60%

Remuneration

The pay structure is the same whether you manage your budget independently or entrust it to one of the algorithms. At the end of the game, your total cumulative balance earned in the ten game rounds is considered. 95 of the originally allocated 100 ECU (10 ECU each in 10 game rounds) will be deducted from your balance. The remaining amount will be exchanged for real money at the ratio of 1 ECU = €1 and paid to you as your remuneration.

[3] In the original instructions, this table was referred to as "Table 2."

Procedure

After reading the instructions and answering the control questions, the first remunerated game round (period 11 of 20) starts on your screen.

At the beginning of each game round, you will see the price development of the Z Share, the development of the influencing factors and the development of the random influence for the last ten game rounds (period 1 to 10), in order to get an idea of the development. In addition, you will always be informed of the current values of the four influencing factors for the respective game round. The value of the random influence, on the other hand, is unknown in advance. Afterwards, you make your decision for the respective game round whether you want to manage your round budget independently or entrust it to Algorithm A or entrust it to Algorithm B.

If you decide to do the investment task on your own, next you will choose whether you want to invest 10 ECU in Z Shares or save them in the particular round.

If you decide to use an algorithm, it will make the decision between investing and saving in your place.

After submitting the decision, you will be informed about the development of the Z Share price in any case, regardless of whether your budget was invested or saved. So, you will receive the full information in any case. The achieved return from the investment in Z Shares or the saved amount will be credited to your balance.

A total of ten rounds will be played. After the experiment is completed, you will receive your remuneration, which is calculated according to the scheme described under "Remuneration."

Remarks

- Please keep quiet during the experiment!
- Do not look at your neighbor's screen!
- Apart from a pen and a pocket calculator, no other aids (smartphones, smartwatches, etc.) are permitted.
- Only use the white sheet of paper provided for your notes.

Appendix B: Test Questions

Question 1: How many rounds of play does this economic experiment involve?

a) 5.
b) 10. (correct)
c) 15.

Question 2 (condition 1): What alternatives do you have in each round?

a) I must perform the investment task independently.
b) I can perform the investment task independently or delegate it to a financial expert.
c) I can perform the investment task independently or delegate it to a robo-advisor (algorithm). (correct)

Question 2 (condition 2): What alternatives do you have in each round?

a) I must perform the investment task independently.
b) I can perform the investment task independently or delegate it to a financial expert.
c) I can perform the investment task independently or delegate it to one of two robo-advisors (algorithms). (correct)

Question 3: Which influencing factors have a positive effect on the price of the Z Share?

a) Influencing factors A, B, and C.
b) Influencing factors A, B, and D. (correct)
c) Influencing factors A, C, and D.

Question 4: How is your remuneration calculated?

a) At the end of the game, 100 ECU will be deducted from my balance. The remaining amount will be exchanged for real money at the ratio of 1 ECU = €0.10.
b) At the end of the game, 100 ECU will be deducted from my balance. The remaining amount will be exchanged for real money at the ratio of 1 ECU = €1.00.
c) At the end of the game, 95 ECU will be deducted from my balance. The remaining amount will be exchanged for real money at the ratio of 1 ECU = €1.00. (correct)

Appendix C: Screen Design

Figure A-1: Screen design in condition 1 (control)

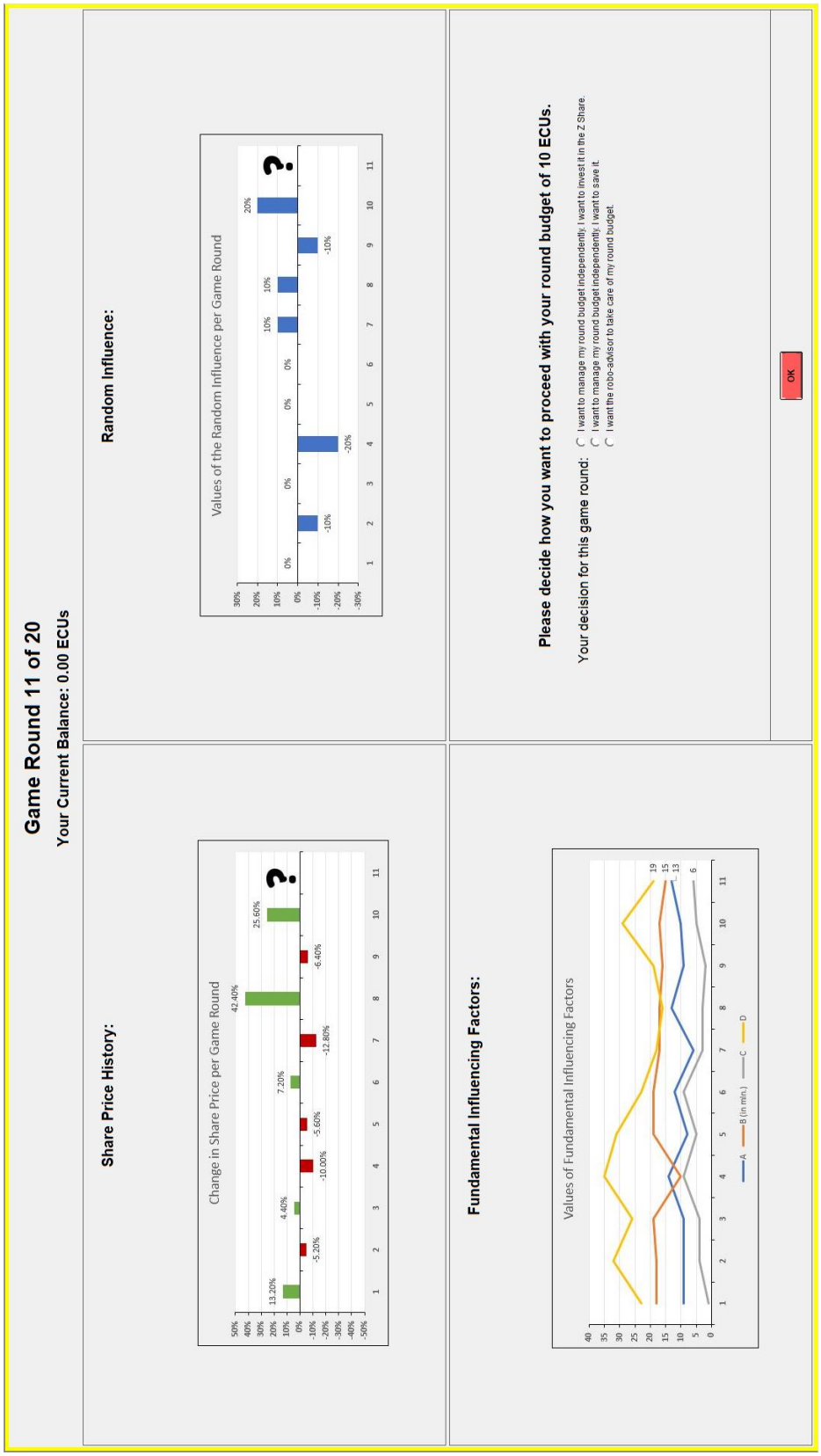
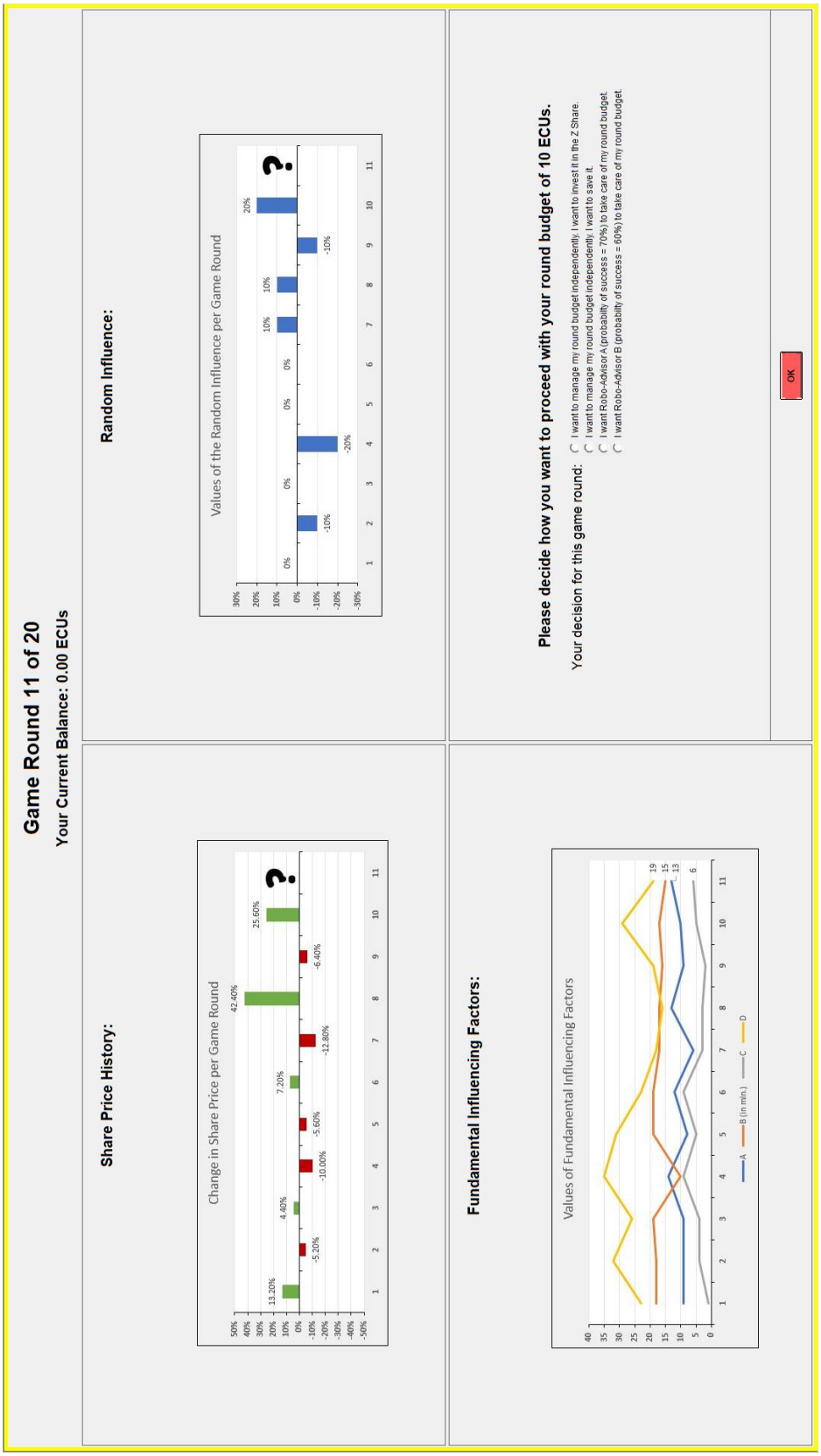


Figure A-2: Screen design in condition 2 (decoy)



Chapter VI

Algorithm Aversion as an Obstacle in the Establishment of Robo Advisors

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Published in:
Journal of Risk and Financial Management, 15(8), 353.

Also published in:
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 22-01, Wolfsburg, July 2021. (Working Paper)

Abstract

Within the framework of a laboratory experiment, we examine to what extent algorithm aversion acts as an obstacle in the establishment of robo advisors. The subjects have to complete diversification tasks. They can either do this themselves or they can delegate them to a robo advisor. The robo advisor evaluates all the relevant data and always makes the decision which leads to the highest expected value for the subject's payment. Although the high level of efficiency of the robo advisor is clear to see, the subjects only entrust their decisions to the robo advisor in around 40% of cases. In this way they reduce their success and their payment. Many subjects orientate themselves towards the 1/n-heuristic, which also contributes to their sub-optimal decisions. As long as the subjects have to make decisions for others, they noticeably make a greater effort and are also more successful than when they make decisions for themselves. However, this does not have an effect on their acceptance of robo advisors. Even when they make decisions on behalf of others, the robo advisor is only consulted in around 40% of cases. This tendency towards algorithm aversion among subjects is an obstacle to the broader establishment of robo advisors.

Keywords

Algorithm aversion, robo advisors, decisions for others, portfolio choice, diversification, behavioral finance, experiments.

JEL Classification

D81, D84, D91, G11, G21, G41, O31, O33.

Journal of
Risk and Financial Management

MDPI

Article

Algorithm Aversion as an Obstacle in the Establishment of Robo Advisors

Ibrahim Filiz [1,*], Jan René Judek [1], Marco Lorenz [2] and Markus Spiwoks [1]

[1] Faculty of Business, Ostfalia University of Applied Sciences, Siegfried-Ehlers-Str. 1, D-38440 Wolfsburg, Germany
[2] Faculty of Economic Sciences, Georg August University Göttingen, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany
* Correspondence: ibrahim.filiz@ostfalia.de

Abstract: Within the framework of a laboratory experiment, we examine to what extent algorithm aversion acts as an obstacle in the establishment of robo advisors. The subjects had to complete diversification tasks. They could either do this themselves or they could delegate them to a robo advisor. The robo advisor evaluated all the relevant data and always made the decision which led to the highest expected value for the subjects' payment. Although the high level of efficiency in the robo advisor was clear to see, the subjects only entrusted their decisions to the robo advisor in around 40% of cases. In this way, they reduced their success and their payment. Many subjects orientated themselves towards the 1/n-heuristic, which also contributed to their suboptimal decisions. As long as the subjects had to make decisions for others, they noticeably made a greater effort and were also more successful than when they made decisions for themselves. However, this did not have an effect on their acceptance of robo advisors. Even when they made decisions on behalf of others, the robo advisor was only consulted in around 40% of cases. This tendency towards algorithm aversion among subjects is an obstacle to the broader establishment of robo advisors.

Keywords: algorithm aversion; robo advisors; decisions for others; portfolio choice; diversification; behavioural finance; experiments

JEL Classification: D81; D84; D91; G11; G21; G41; O31; O33

Citation: Filiz, Ibrahim, Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2022. Algorithm Aversion as an Obstacle in the Establishment of Robo Advisors. *Journal of Risk and Financial Management* 15: 353. https://doi.org/10.3390/jrfm15080353

Academic Editor: Mario Maggi

Received: 5 July 2022
Accepted: 4 August 2022
Published: 8 August 2022

1. Introduction

The traditional portfolio management business is demanding in terms of human resources and therefore comparatively expensive. Wealthy private customers have, however, become more price-sensitive since the establishment of low-cost investment opportunities such as exchange-traded funds (ETFs) in recent decades. Many banks are thus trying to find low-cost alternatives, particularly for the support of customers with smaller and medium-sized assets. The increased use of automated processes in portfolio management offers considerable scope for cost reduction. Many banks thus offer robo advisors (see, for example, Rühr et al. 2019a; Jung et al. 2018; Singh and Kaur 2017). Robo advisors are algorithms which are specialised in making investment decisions for customers and processing them. Using new technologies such as artificial neural networks, robo advisors are becoming increasingly more powerful and can potentially maximise clients' returns (Méndez-Suárez et al. 2019).

However, many customers have reservations about interacting with automated processes (robo advisors), although the latter are often remarkably effective (see, for example, Rossi and Utkus 2020; Bhatia et al. 2020; D'Acunto et al. 2019; Beketov et al. 2018; Uhl and Rohner 2018). So-called algorithm aversion is thus a significant problem for the banking sector.

Algorithm aversion particularly occurs when algorithms have to deal with stochastic processes. This is undoubtedly the case with robo advisors. Even when the algorithm makes very good investment decisions, it will—given the stochastic nature of financial market trends—never be able to always make perfect investment decisions. Dietvorst et al. (2015) showed that the tolerance of occasional errors by algorithms is much lower than the tolerance shown regarding occasional poor decisions which one has taken oneself or are made by an expert. We speak of algorithm aversion when subjects decline the use of an algorithm even though it is clearly recognisable that their own decisions or those of experts are by no means more successful (for the usual definitions, see, for example, Filiz et al. 2021a). There is a considerable amount of research results available on measures which can mitigate algorithm aversion (see, for example, Hinsen et al. 2022; Filiz et al. 2021b; Gubaydullina et al. 2021; Kim et al. 2021; Jung and Seiter 2021; Castelo et al. 2019; Dietvorst et al. 2018; Taylor 2017).

The efficiency of robo advisors is due—among other things—to the fact that they can make meaningful diversification decisions effortlessly. By contrast, investors often find it difficult to determine the expected earnings and the risk (variance) of alternative investments and to take into account the correlations of different investment opportunities in an appropriate way (see, for example, Ungeheuer and Weber 2021; Cornil et al. 2019; Enke and Zimmermann 2019; Gubaydullina and Spiwoks 2015; Eyster and Weizsäcker 2011; Kallir and Sonsino 2009; Hedesstrom et al. 2006). This is why in practice many portfolios prove to be under-diversified or diversified in unsuitable ways (see, for example, Gomes et al. 2021; Chu et al. 2017; Dimmock et al. 2016; Anderson 2013; Hibbert et al. 2012; Goetzmann and Kumar 2008; Meulbroek 2005; Polkovnichenko 2005; Huberman and Sengmueller 2004; Agnew et al. 2003; Guiso et al. 2002; Benartzi 2001; Benartzi and Thaler 2001; Barber and Odean 2000; Bode et al. 1994; Blume and Friend 1975; Lease et al. 1974).

We build on studies that have examined what influences the willingness to use a robo advisor. Alemanni et al. (2020) showed that the willingness to follow a robo advisor is lower when the robo advisor suggests a portfolio change. If the current portfolio is to be retained, the willingness to use is similar to advice from human advisors (Alemanni et al. 2020). In a questionnaire-based study, von Walter et al. (2021) found that consumers who believe artificial intelligence is better than human intelligence are more likely to accept advice from a robo advisor. Hodge et al. (2021) showed that subjects follow advice from a robo advisor without a name more closely than advice from a robo advisor that has been given a name. Robo advisors with names tend to be more popular for simple tasks than for complex ones (Hodge et al. 2021). The age of the decision-maker may also be a factor: Robillard (2018) argued that millennials may rely more heavily on robo advisors because this generation has lower trust in fellow humans than other generations.

Users' risk attitudes and attitudes toward automated processes also influence robo advisors and their investment decisions. Robo advisors can identify different risk profiles of their users, although there are large differences in risk preferences within the same investor type group (Boreiko and Massarotti 2020). User preferences, however, have different effects on the perception of and intention to use robo advisors. For financial investments, a higher perceived level of automation leads to higher performance expectations and higher user control leads to lower perceived risk (Rühr et al. 2019b). Since robo advisors should take user preferences into account to increase their usage intent, a performance-control dilemma arises that needs to be mitigated (Rühr 2020).

Another important aspect seems to be the transfer of responsibility to the robo advisor. Niszczota and Kaszás (2020) discovered that moral investment decisions are rather delegated to humans than to robo advisors. On the other hand, Back et al. (2021) showed that subjects feel better in cases of loss if they have delegated some of the responsibility to the robo advisor. For tasks outside the world of finance, it has already been shown that punishment by third parties can be significantly lower if errors are committed by an algorithm rather than by a human (Feier et al. 2022). The idea that, under certain

circumstances, subjects may be happy to hand over responsibility for possible future mistakes to a robo advisor is remarkable, and we explore it in more detail in the present study.

We carried out an economic experiment in which the subjects had to make four investment decisions. They could choose between different investment alternatives in each of the four cases. They were informed of the possible returns, the probability that these returns would materialise, and the correlations of the different investment opportunities. The subjects could either make their own diversification decisions or entrust the task to a robo advisor. The subjects knew that the robo advisor took all of the relevant data into account (the expected value of the returns, the probability that the returns would materialise, and the correlation coefficients of the return development of the different investments), evaluated them optimally, and took them into account in its investment decisions. However, the subjects were also aware of the fact that the robo advisor could not know which random event will occur next. The subjects received the risk-adjusted return of their investment decisions as payment. This had the advantage that the subjects' risk preferences had no meaning for the assessment of the investment alternatives. We examine whether algorithm aversion occurs in this context and whether this can lead to a reduction in risk-adjusted returns. In this context, adding to previous research, we also consider whether algorithm aversion is less pronounced when a person has to make decisions for others.

Some empirical research findings indicate that when making decisions for others a change in the willingness of subjects to take risks can come into play (see, for example, Andersson et al. 2022; Eriksen et al. 2020; Vieider et al. 2016; Pahlke et al. 2015; Füllbrunn and Luhan 2015; Bolton et al. 2015; Pahlke et al. 2012; Chakravarty et al. 2011; Charness and Jackson 2009; Reynolds et al. 2009). This is particularly true when the person for whom a decision is being made is actually present (Polman 2012). Later on, the persons for whom a decision is made may demand that the decision-maker justify their choices. If this is known in advance, it can lead to particular care on the part of the decision-maker. If the decision is delegated to an algorithm, however, the decision-makers do not have to justify their choices. This could possibly contribute towards a reduction in algorithm aversion.

This study examines the circumstances under which robo advisors can become an important complementary tool in wealth management. In addition to the performance of robo advisors in the meaningful diversification of investment alternatives, the reluctance of subjects to use automated processes (in this case: robo advisors) is the focus of attention. Measures to dampen algorithm aversion are of considerable interest. In this context, we raise the question of whether people are more likely to use a robo advisor if the consequences of the robo advisor's decision affect third parties. Exploring this question can help reduce hurdles to establish robo advisors. It also contributes to our understanding of the relatively new research field of algorithm aversion. The rest of this research paper is organised as follows: In Section 2, the experimental design is explained. Section 3 deals with the elaboration of the hypotheses. In Sections 4 and 5, the results of the economic experiments are presented and discussed in the context of the existing literature. To wrap up this study, Section 6 provides a summary of the key findings and a conclusion.

2. Experimental Design

In order to answer the research question, an economic experiment with two treatments was carried out between 20 and 28 April 2022 in the Ostfalia Laboratory of Experimental Economic Research (OLEW) at Ostfalia University of Applied Sciences in Wolfsburg. A total of 160 students of the Ostfalia University of Applied Sciences took part in the experiment. Of these, 112 subjects (70%) were male and 48 subjects (30%) were female. Of the 160 participants, 98 subjects (61.25%) studied at the Faculty of Economics and Business, 38 subjects (23.75%) at the Faculty of Vehicle Technology, and 24 subjects (15%) at other faculties. Their average age was 23.6 years.

In each treatment, subjects have to make four investment decisions (tasks 1–4) whose success directly affects them (or others). However, the subjects do not profit from gains in the share prices—they only profit (once) from the dividend payments of the shares in 2022. The subjects can either make their own diversification decisions or entrust the task to a robo advisor. In the treatment entitled 'Self' the subjects make a diversification decision for their own portfolio and receive the payment themselves. In the treatment entitled 'Representative' the subjects make a diversification decision for another participant's portfolio and the other participant in the session receives the payment which has been obtained. In the treatment 'Representative', after the payment has been made, the subjects are informed about who is responsible for which payment.

Let us assume, for example, that subject C receives the payment achieved by subject B and vice versa (Figure 1). After the experiment, subject B could demand in a personal conversation that subject C justifies his or her decisions. Moreover, subject C could also demand that subject B justifies their decisions. All of the subjects who participate in the treatment 'Representative' are informed about this at the beginning of the experiment.

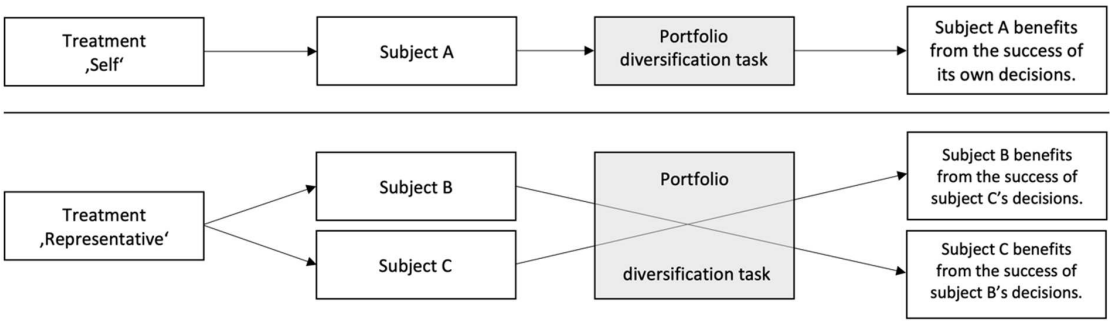


Figure 1. The treatment 'Self' and the treatment 'Representative'.

In the first task, there are two shares to choose from: share Y and share Z. The dividend payments of both companies are independent random processes with two possible configurations: 8 experimental currency units (ECU) and ECU 0. The probability of each of these occurring is 50%. The expected values of the dividend payments are thus ECU 4 each. The dividend payments of the two shares are wholly uncorrelated (correlation coefficient = 0). Table 1 shows the level of the dividend payments of the two shares in the past ten years. In this task, as well as in all other tasks in the experiment, subjects are given the dividend payments for the years 2012 to 2021. The dividend payments for 2022, which are relevant for their payoff, are still unknown, which is illustrated by the question mark.

Table 1. History of the random events of the dividend payments in task 1.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share Y	ECU 8	ECU 0	ECU 8	ECU 8	ECU 8	ECU 0	ECU 8	ECU 0	ECU 0	ECU 0	?
Share Z	ECU 0	ECU 0	ECU 8	ECU 0	ECU 8	ECU 8	ECU 0	ECU 0	ECU 8	ECU 8	?

The subjects are allowed to compile a portfolio consisting of two shares. They can thus choose two Y shares, two Z shares, or one Y share and one Z share. As payment, they receive the risk-adjusted dividends for 2022. A risk-adjusted dividend is equivalent to the dividend payment divided by the variance of the dividend payments of the chosen portfolio. The task thus consists of achieving the highest possible dividends with the lowest possible risk (low variance). The total of all risk-adjusted dividends (in ECU) which are obtained via portfolio decisions is multiplied by five at the end and then paid in euros.

As the subjects do not know the next random events for the dividend payments of share Y and share Z, it makes sense for them to orientate themselves towards the expected values and the variances of the three possible portfolios (see Table 2).

Table 2. Expected values and variances in task 1.

Possible Portfolios	Expected Value of the Dividend	Variance	Expected Value of the Payment
2 Y shares	ECU 8	64	ECU 0.125 or EUR 0.625
2 Z shares	ECU 8	64	ECU 0.125 or EUR 0.625
1 Y share + 1 Z share	ECU 8	32	ECU 0.25 or EUR 1.25

Rational economic subjects orientate themselves towards the expected values of the payment, i.e., they select the mixed securities portfolio (1 Y share + 1 Z share). This is exactly how the robo advisor works.

All of the subjects have been familiarised with stochastic processes and the calculation of probabilities at school and also at the beginning of their degree programmes. They are aware of the fact that one cannot draw any conclusions about future random occurrences from an independent random event. Nevertheless, the temptation is great to make a forecast on which events will occur in the cases of the two shares in 2022 which is derived from the sequence of favourable and unfavourable dividend payments. People tend to see patterns even where there are definitely none (see, for example, Zielonka 2004; Wärneryd 2001; Gilovich et al. 1985; Roberts 1959). Subjects who have succumbed to the hot hand fallacy (Burns 2001; Gilovich et al. 1985) will tend to choose the portfolio of 2 Z shares. Subjects who believe in the gambler's fallacy (Rogers 1998; Tversky and Kahneman 1971) will prefer the 2 Y shares portfolio. Subjects who think they can predict the next random events will not make use of the robo advisor. Subjects who want to maximise the expected value of their payment can, however, sleep easily if they delegate the decision to the robo advisor because the robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way in order to achieve risk-adjusted dividend payments which are as high as possible. The subjects are informed of this.

The second task is somewhat more complex. Once again, there are two shares to choose from (share X and share Q). Both of the shares can pay a dividend of either ECU 4 or ECU 0. The probability of each of these occurring is 50%. The expected values of the dividend payments are thus ECU 2 each. Once again, they are independent random events. The dividend payments of share X and share Q are completely uncorrelated (correlation coefficient = 0).

Table 3 shows the level of the dividend payments of the two shares in the last 10 years.

Table 3. History of the random events of the dividend payments in task 2.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share X	ECU 0	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 4	ECU 4	?
Share Q	ECU 0	ECU 4	ECU 4	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 4	ECU 0	?

The subjects can compile a portfolio consisting of four shares. They can thus choose four X shares, four Q shares, three X shares and one Q share, three Q shares and one X share, or two X shares and two Q shares. Neither the subjects nor the robo advisor know what the random events (dividend payments for share X and share Q) will be in 2022. A rational subject would orientate themselves towards the expected value of the payment and select the portfolio 2 X shares + 2 Q shares (see Table 4). This is exactly what the robo advisor does.

Table 4. Expected values and variances in task 2.

Possible Portfolios	Expected Value of the Dividend	Variance	Expected Value of the Payment
4 X shares	ECU 8	64	ECU 0.125 or EUR 0.625
4 Q shares	ECU 8	64	ECU 0.125 or EUR 0.625
3 X shares + 1 Q share	ECU 8	40	ECU 0.20 or EUR 1
3 Q shares + 1 X share	ECU 8	40	ECU 0.20 or EUR 1
2 X shares + 2 Q shares	ECU 8	32	ECU 0.25 or EUR 1.25

The third task and the fourth task can no longer be accomplished with a crude diversification strategy such as the 1/n heuristic (see, for example, Fernandes 2013; Baltussen and Post 2011) because these are companies which belong to the same industry sector and whose dividend payments depend on the success of the sector. The dividend payments of the two shares are thus completely positively correlated (correlation coefficient = 1). Table 5 shows the amount of the dividend payments in the past ten years.

Table 5. History of the random events of the dividend payments in task 4.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share M	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 0	ECU 4	?
Share P	ECU 3	ECU 1	ECU 3	ECU 1	ECU 1	ECU 1	ECU 3	ECU 3	ECU 1	ECU 3	?

A phase in which companies in this sector are either successful or are struggling occurs purely coincidentally with a probability of 50%. Previous events thus provide no indication of which random events might occur in the future. The expected value of the dividend payments is thus ECU 2 for both shares. The subjects can compile a portfolio consisting of four shares.

Given that the dividend payments for both shares are 100% positively correlated, a mixture of the two shares does not create any diversification effect. The optimal strategy is to select four P shares because that is the minimum variance portfolio (see Table 6). This is precisely the strategy pursued by the robo advisor.

Table 6. Expected values and variances in task 4.

Possible Portfolios	Expected Value of the Dividend	Variance	Expected Value of the Payment
4 M shares	ECU 8	64	ECU 0.125 or EUR 0.625
4 P shares	ECU 8	16	ECU 0.50 or EUR 2.50
3 M shares + 1 P share	ECU 8	49	ECU 0.165 or EUR 0.825
3 P shares + 1 M share	ECU 8	25	ECU 0.32 or EUR 1.60
2 M shares + 2 P shares	ECU 8	36	ECU 0.225 or EUR 1.125

The experiment proceeds as follows: First, the subjects read the instructions and answer the control questions (see Appendices A and B). Afterwards, they make the four portfolio decisions of tasks 1 to 4 either with the help of the robo advisor or independently (see Appendix C). For each of the four tasks, the subjects can decide again whether they want to delegate the task to the robo advisor or whether they want to choose a portfolio composition themselves. Only after the four tasks have been completed is it revealed which random events have occurred in this session and to which compensation the subjects have progressed. The payment is then made in cash.

3. Hypotheses

The most meaningful strategy is to delegate all four tasks to the robo advisor. The robo advisor always makes the most meaningful decisions. It always selects the portfolio composition which maximises the expected value of the payment in euros. It would actually be possible to work out this optimal decision oneself. However, the amount of effort required to do so is considerable. The subjects can make mistakes when calculating the expected payment amount. The robo advisor, on the other hand, always evaluates all of the relevant data in an optimal way and always makes the decision which maximises the expected value of the payment. Nevertheless, it has to be expected that some subjects will have reservations about using a robo advisor. The wide variety of previous findings on the occurrence of algorithm aversion make this highly likely (Mahmud et al. 2022; Kawaguchi 2021; Burton et al. 2020; Castelo et al. 2019, Prahl and Van Swol 2017).

Hypothesis 1. *Not all of the subjects will trust the robo advisor (algorithm), although it is not possible for them to make a better decision. This means that algorithm aversion will occur.*

Null Hypothesis 1. *All of the subjects will trust the robo advisor (algorithm). This means that algorithm aversion will not occur.*

If the subjects are wary of using the robo advisor (algorithm aversion), this may well lead—on average—to a reduction in the payment they obtain. Algorithm aversion will presumably cause a loss in potential earnings.

Hypothesis 2. *The more frequently the subjects delegate their decision to the robo advisor, the higher their payments will be.*

Null Hypothesis 2. *The frequency with which the subjects delegate their decisions to the robo advisor does not have a positive influence on their payment.*

Among the subjects, there will presumably be some who pursue a crude diversification strategy (1/n-heuristic; see, for example, Fernandes 2013; Morrin et al. 2012; Baltussen and Post 2011; Huberman and Jiang 2006; Benartzi and Thaler 2001). This strategy can lead to success in tasks 1 and 2. In tasks 3 and 4, on the other hand, it cannot lead to success. For an optimal solution of tasks 3 and 4, it is necessary to also take into account the correlation coefficients alongside the expected values of the dividends.

Hypothesis 3. *Subjects who do not deploy the algorithm partly neglect the correlations, and in the cases of tasks 3 and 4 they find the optimal solution significantly less often than in tasks 1 and 2.*

Null Hypothesis 3. *Subjects who do not deploy the algorithm do not neglect the correlations, and in the cases of tasks 3 and 4 they do not find the optimal solution significantly less often than in tasks 1 and 2.*

On the basis of the existing research on decision making for others (see, for example, Pahlke et al. 2015; Polman 2012; Pahlke et al. 2012; Charness and Jackson 2009; Reynolds et al. 2009) we presume that the subjects who make decisions for others (the treatment 'Representative') consider their decisions more carefully and try harder to make meaningful decisions. After all, the persons for whom the decisions are being made are actually present. At the end of the experiment, who decided for whom and what the results were is announced. All of the subjects in the treatment 'Representative' are aware of this. In other words, they have to expect that they will need to justify their decisions. The subjects in the treatment 'Self', on the other hand, are only responsible for themselves. They need not fear that someone will demand that they justify their decisions. We therefore presume that algorithm aversion will occur less frequently in the treatment 'Representative' than

in the treatment 'Self'. In addition, we presume that those persons in the treatment 'Representative' who do not want to trust the robo advisor—for whatever reason—will make a greater effort to select meaningfully diversified portfolios.

Hypothesis 4. *The solution of the tasks is delegated to the robo advisor significantly more often in the treatment 'Representative' than in the treatment 'Self'.*

Null Hypothesis 4. *The solution of the tasks is not delegated to the robo advisor significantly more often in the treatment 'Representative' than in the treatment 'Self'.*

Hypothesis 5. *Those persons who do not want to trust the robo advisor will choose the optimal portfolio structure significantly more often in the treatment 'Representative' than in the treatment 'Self'.*

Null Hypothesis 5. *Those persons who do not want to trust the robo advisor will not choose the optimal portfolio structure significantly more often in the treatment 'Representative' than in the treatment 'Self'.*

The general research question of this study is: Can robo advisors become useful complementary tools in the modern wealth management business? In order to explore our research question, we assume that a robo advisor cannot forecast future capital market developments without errors. However, a robo advisor can effortlessly make meaningful diversification decisions. This leads to the question if economic agents need a robo advisor in order to achieve good diversification decisions with certainty. In four very clear decision situations where shares are assembled into a portfolio, optimal decisions can easily be made. However, the facts (expected value, the dispersion of events around expected value, and the correlation of the events of different shares) are neglected or misinterpreted by many economic agents. Therefore, a greater willingness to delegate the decision to the robo advisor presumably leads to greater investment success or higher compensation (Hypothesis 2). The fact that is most often neglected is probably the correlation of the returns of different shares (Hypothesis 3).

Although the subjects know that the robo advisor optimally evaluates all relevant information and makes the best possible diversification decision in each case, experience has shown that many economic subjects are reluctant to entrust themselves to an algorithm—in this case a robo advisor (Hypothesis 1). Thus, if robo advisors are to be successfully established, measures to mitigate algorithm aversion have to be considered. One possible measure would be to place the decision to use a robo advisor in the context of decision for others. After all, investment decisions are not only important for the wealthy person but also for his or her family, especially children and grandchildren. Thus, in the case of decision for others, the willingness to use the robo advisor might increase (Hypothesis 4) because economic agents might try harder to make a meaningful decision when making decisions that (also) affect others (Hypothesis 5).

4. Results

Of the 160 participants, 80 subjects played the treatment 'Self' and 80 played the treatment 'Representative'. The experiment was carried out using z-Tree (Fischbacher 2007). The time needed for reading the instructions of the experiment (Appendix A), answering the test questions (Appendix B), and carrying out the four tasks took 15 min on average. An average payment of EUR 6.89 seemed very attractive for the amount of time required. It was intended to be sufficient incentive for meaningful economic decisions, and the subjects did actually give the impression of being concentrated and motivated.

In the first instance, it could be seen that algorithm aversion occurred to a considerable extent. Although it was clear to all of the participants that using the algorithm (robo advisor) definitely led to the best possible decisions, the robo advisor was deployed in

less than half of the cases. A total of 160 subjects had to make four decisions each. This was a total of 640 decisions. The subjects decided to delegate the task to the robo advisor in only 258 cases (40.31%). In 382 cases (59.69%), the subjects refrained from using the algorithm (Figure 2). The reason why this is so remarkable is that all of the subjects knew that the robo advisor evaluated all of the relevant data in an optimal way and therefore always made the best possible decision.

An average subject relied on the algorithm in only 1.612 out of 4 rounds. The *t*-test shows in all clarity that Null Hypothesis 1 has to be rejected (*p*-value = 0.000). The Z-test supports that only very few subjects (36 out of 160) consistently followed the rational strategy and relied on the algorithm in all rounds of the experiment (*p*-value = 0.000). Algorithm aversion thus obviously occurred to a considerable extent (59.69% of all decisions).

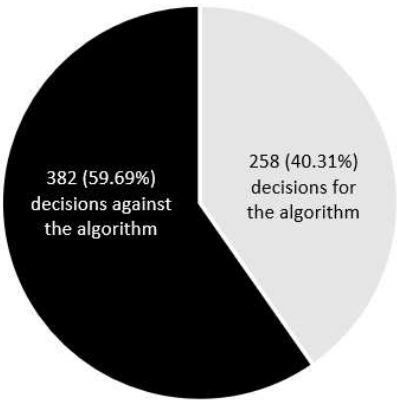


Figure 2. Decisions for and against the algorithm (robo advisor).

It is of particular interest whether this tendency towards algorithm aversion really led to a smaller number of optimal diversification decisions and whether the payments were lower than would have been the case when the subjects had consistently trusted the robo advisor. After all, one cannot simply presume that the decisions of the subjects who did not always use the robo advisor were really less successful.

A total of 53 subjects did not delegate their decision to the robo advisor a single time. In 89 out of 212 decisions (41.98%), these subjects selected optimal portfolios. On average, they achieved an expected payment value of EUR 6.36. How much the actual payment was also depended on the specific random events (dividend payments). Here, there was an average payment of EUR 6.67 (Table 7).

A total of 34 subjects delegated all four of their decisions to the robo advisor. As was to be expected, in 136 out of 136 decisions (100%), the optimal portfolios were chosen. The subjects achieved an expected payment value of EUR 8.13. The specific random events (dividend payments) led to an average payment of EUR 7.53 (Table 7).

Table 7. Average success in relation to the extent of algorithm aversion.

Number of Times the Algorithm Was Chosen	Number of Subjects	Optimal Portfolios	Expected Value of the Payment in Euros	Actual Payment in Euros
0	53	89 (41.98%)	EUR 6.36	EUR 6.67
1	39	71 (45.51%)	EUR 6.20	EUR 6.49
2	19	51 (67.11%)	EUR 7.23	EUR 6.78
3	15	48 (80.00%)	EUR 7.37	EUR 7.46
4	34	136 (100%)	EUR 8.13	EUR 7.53

Figure 3 shows clearly that the more frequently the subjects delegated their decision to the robo advisor, the more successful they were. The subjects who did not put their faith in the robo advisor a single time achieved an average of only 1.68 optimal portfolios. The subjects who used the robo advisor to solve all four tasks made 4.00 optimal decisions. The F-test confirms: the more frequently the robo advisor was used, the more optimal portfolios were compiled (thick grey line, left scale, *p*-value = 0.000), the higher the expected value of payment (dashed black line, right scale, *p*-value = 0.000), and the higher the actual payment (continuous black line, right scale, *p*-value = 0.000).

The stronger the effect of algorithm aversion, the less successful the subjects were. Null Hypothesis 2 thus has to be discarded.

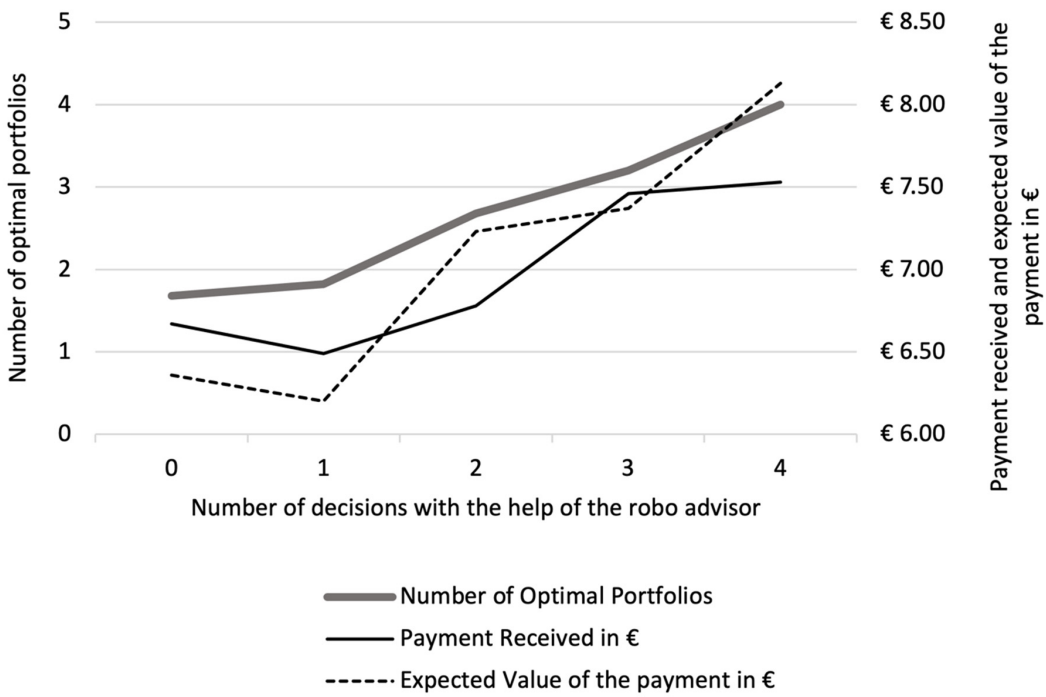


Figure 3. Average success in relation to the extent of algorithm aversion.

Now let us look at the success of the decisions which were not delegated to the robo advisor. Tasks 1 and 2 can be solved well with the simple understanding of the diversification of the 1/n heuristic. In tasks 3 and 4, however, it is absolutely necessary to take the correlations between the dividend payments of the two shares into account and to understand the variances of the dividend payments of the two shares. Among the decisions which are not delegated to the robo advisor, a clear difference can indeed be seen between the success rate in tasks 1 and 2 on the one hand and the success rates in tasks 3 and 4 on the other. In tasks 1 and 2, 81 out of 182 decisions (44.51%) led to optimal portfolios. In tasks 3 and 4, on the other hand, only 56 out of 200 decisions (28%) led to optimal portfolios which maximised the expected value of the payment. In the chi square test, this difference proves to be significant (*p*-value = 0.001). Null Hypothesis 3 thus has to be rejected (Figure 4).

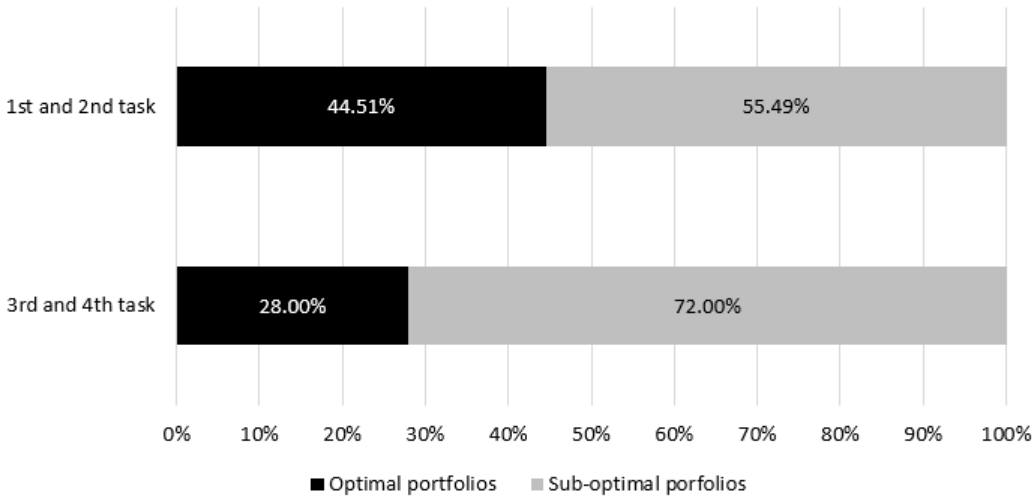


Figure 4. Percentage share of optimal portfolios according to tasks.

In a comparison of the two treatments 'Self' and 'Representative', no noteworthy differences with regard to use of the robo advisor can be seen. In the treatment 'Self', 131 of out 320 decisions (40.94%) were delegated to the robo advisor. In the treatment 'Representative', 127 out of 320 decisions (39.69%) were delegated to the robo advisor (Table 8 and Figure 5). This is only a very small difference. It proves to be insignificant both in the Wilcoxon rank sum test (p-value = 0.752) as well as in the chi square test (p-value = 0.747). Null Hypothesis 4 can therefore not be rejected.

Table 8. Influence of the treatments on algorithm aversion.

Treatment	Robo Advisor	Own Decision	Total
'Self'	131	189	320
'Representative'	127	193	320

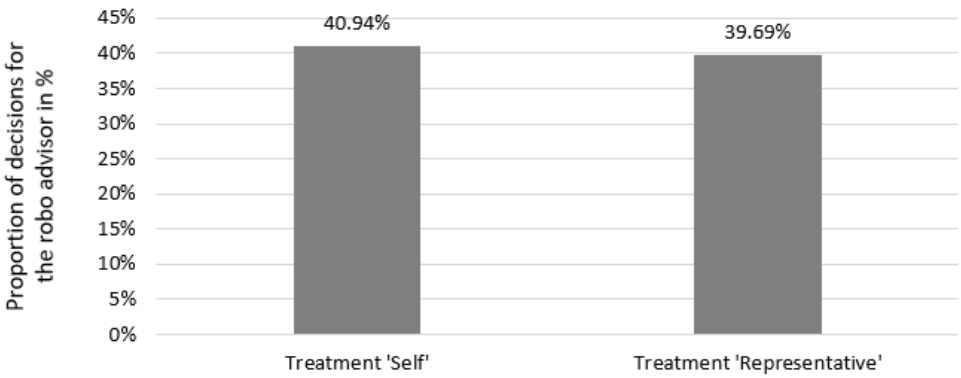


Figure 5. Acceptance of the robo advisor according to treatments.

This is a surprising result. The subjects in the treatment 'Representative' could have easily transferred their responsibility for the payment of another person to the robo advisor. Given that the robo advisor was known for the fact that it always made optimal decisions, nobody needed to be afraid of being criticised. However, a large part of the subjects obviously had such far-reaching reservations regarding the deployment of a robo advisor

that they did not want to take this route. We thus have to come to the conclusion that algorithm aversion occurs frequently and is by no means easy to overcome.

However, it is noticeable that it does make a difference whether one makes decisions for oneself or for others. The subjects in the treatment 'Representative' really did make a greater effort to make meaningful decisions. This can be seen in the decisions they made without using the robo advisor. In 57 out of 189 decisions (30.16%) the subjects in the treatment 'Self' succeeded in building optimal portfolios (portfolios with the highest expectation value for the payment in euros). In 80 out of 193 decisions (41.45%), the subjects in the treatment 'Representative' succeeded in building optimal portfolios (portfolios with the highest expectation value for the payment in euros) (Table 9 and Figure 6). This difference turns out to be statistically significant in the chi square test (p-value = 0.021).

Table 9. Success of portfolio decisions without the robo advisor according to treatments.

Treatment	Number of Subjects	Number of Optimal Portfolios without the Robo Advisor	Number of Suboptimal Portfolios without the Robo Advisor	Number of Decisions Made by the Robo Advisor	Total
'Self'	80	57	132	131	320
'Representative'	80	80	113	127	320

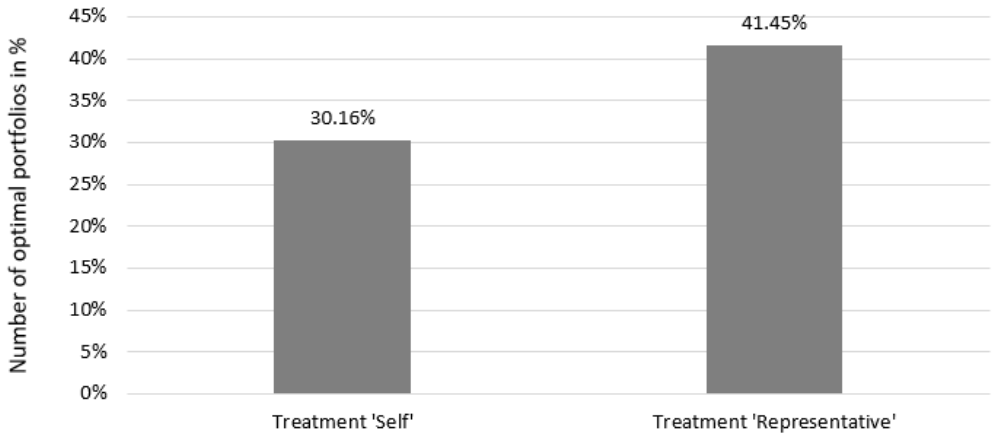


Figure 6. Success of the portfolio decisions without the robo advisor according to treatments.

A clear difference between the two treatments can definitely be seen. The subjects behaved differently depending on whether they were deciding for themselves or for others. They obviously acted less impulsively in the treatment 'Representative' and weighed up more precisely which portfolio composition would presumably lead to the largest payment. However, this effort to make meaningful decisions did not lead to a greater acceptance of robo advisors. The subjects' reservations about using an algorithm were obviously stronger than their wishes to make decisions for others with particular care.

5. Discussion

Our results contribute to the academic debate in three ways. First, it has been shown that many subjects have massive reservations about robo advisors despite their obvious advantages. In our study, robo advisors consistently outperformed subjects. Still, most subjects chose not to use them. Although robo advisors have enormous potential and perform significantly better on average, they seemed to be very unpopular among subjects. This is in line with previous studies, which also found that algorithm aversion in

particular can be a hurdle in establishing robo advisors (Hodge et al. 2021; Alemanni et al. 2020; Niszczota and Kaszás 2020).

Second, our research confirms that algorithm aversion is a serious barrier to the diffusion of innovative business fields in general. In this respect, we may also be facing a societal problem. Already today, the use of algorithms clearly provides humans with more powerful options for solving problems. Yet, decision-makers refuse to use them. Instead, they perform tasks themselves, leading to higher costs and poorer results. It therefore remains an important task of research, especially with regard to cognitive biases and heuristics, to further explore the background of algorithm aversion in order to contribute to the progress of society.

Third, it turns out that it makes little difference to the extent of algorithm aversion who has to bear the consequences (oneself or third parties). Research by Back et al. (2021) suggests that one reason to consult a robo advisor might be that it feels like relinquishing some of the responsibility for unpleasant tasks and potential mistakes. However, this assumption was not confirmed in our study. If subjects made decisions for others who may have demanded a justification for possible mistakes, the robo advisor was nevertheless just as unpopular.

To save taxes, many wealthy private clients transfer part of their assets to their children while they are still minors. These assets also need to be managed. The parents now have to decide on behalf of their children how this should be accomplished. If algorithm aversion were less prominent in decisions for others, this could be a starting point to resolve or at least mitigate the bias against robo advisors. However, no evidence for this has emerged. Algorithm aversion is reflected to the same extent in the decisions that economic agents make for themselves and in the decisions that they make for others.

Of course, there are also some limitations that may affect the validity of our results for practical applications. First, it should be mentioned that the results were obtained in the context of financial decisions with robo advisors. Financial decisions are influenced by a variety of factors, such as financial literacy or experience. Algorithm aversion is far from being the only influencing factor. It may therefore be worthwhile to revisit our research question in relation to other areas of use for algorithms.

Moreover, robo advisors from reputable banks go through a detailed accreditation process. In this process, independent experts verify, for example, whether the robo advisors take appropriate measures to hedge risks and also make decisions that are justifiable from an ethical point of view. Accreditation is thus a tool that can increase user confidence. However, it cannot be replicated in the same way in an economic laboratory experiment.

Finally, when making decisions on behalf of others, it may always make a difference what one's relationship is to the person who has to bear the consequences. We conducted a laboratory experiment at our research institution. Usually, students go there together with fellow students whom they know from classes. Sometimes students also come alone. As such, the consequences of the decision in the treatment 'Representative' were largely borne either by complete strangers or loose acquaintances. It must be left to future research efforts to see if a different outcome emerges when we decide, for example, on behalf of loved ones.

6. Conclusions

Robo advisors are algorithms which can automatically make investment decisions for asset management customers. Given the increased price sensitivity of wealthy private clients, robo advisors are one way to offer solid portfolio management decisions at a low cost. However, customers have considerable reservations about algorithms, even when they are very efficient systems. This phenomenon, which is known as algorithm aversion, is considered in more detail in this study.

In a laboratory experiment, subjects made a total of four portfolio decisions. They could either try to determine the optimal portfolio composition in each case themselves or they could delegate the task to a robo advisor. The robo advisor took all the relevant

information into account in an optimal way and always chose the portfolio composition which led to the highest expected value of the payment in euros. The subjects were familiar with the qualities of the robo advisor. Nevertheless, they only used it in around 40% of all cases. In around 60% of all decision making situations, the subjects trusted in their own judgement, although it must have been clear to them that they were not able to make better decisions than the robo advisor. Algorithm aversion thus occurred to a great extent.

The actual success rate of the subjects who did not put their faith in the robo advisor was indeed lower than that of the robo advisor. This applied to the average number of optimal portfolio compositions, to the average expected values of their payment in euros, and also with regard to the actually obtained payment in euros. It is crystal clear that the more frequently the subjects delegated their decisions to the robo advisor, the greater their success. With their aversion towards the algorithm, the subjects were recognisably damaging themselves.

The subjects had particular difficulties when trying to take into account the correlation between the different investments. Tasks which could be solved with the simple diversification strategy of the 1/n heuristic (tasks 1 and 2) were dealt with successfully significantly more often than tasks which could not be suitably dealt with using the 1/n heuristic (tasks 3 and 4).

Ultimately, it became clear that subjects who had to make decisions for others approached the task in a more careful and concentrated way. Among the decisions which were not made by the robo advisor, there were significantly more optimal portfolios within the subjects who made decisions for others than among those who decided for themselves. However, this did not have an effect on algorithm aversion. Regardless of whether the subjects decided for themselves or for others, a readiness to delegate the decision to the robo advisor could only be seen in around 40% of decisions.

To summarise, the following can be stated: The deployment of robo advisors can, under certain circumstances, be a low-cost and very efficient alternative to traditional asset management. However, algorithm aversion hinders the establishment of the business which could be had with robo advisors.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jrfm15080353/s1.

Author Contributions: Conceptualization, M.S.; Software, I.F., J.R.J., and M.L.; Validation, I.F., J.R.J., M.L., and M.S.; Formal analysis, I.F., J.R.J., M.L., and M.S.; Data curation, I.F., J.R.J., M.L., and M.S.; Writing—original draft preparation, J.R.J., M.L., and M.S.; Writing—review and editing, I.F., J.R.J., M.L., and M.S.; Visualization, J.R.J., and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Our data can be accessed at Supplementary Materials.

Acknowledgments: The authors thank the editor and the anonymous reviewers for their constructive comments and useful suggestions, which were very helpful to enhance the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Instructions for the Experiment

Appendix A.1. Instructions (Treatment 'Self')

You have the task of creating portfolios of shares. A portfolio of shares is a compilation of several shares.

The development of the share prices is of no concern to you, because you profit only once from the dividend payments of the shares in 2022. The dividend is the distribution of profits of a stock exchange-listed company to its shareholders.

You will receive information about how the dividend payments might turn out, and about the probabilities of different amounts of dividend. In addition, you will be shown how the dividends of the shares have developed over the last ten years.

You are paid the risk-adjusted dividend. A risk-adjusted dividend is the dividend payment divided by the variance of the dividend payments of the selected portfolio. Your task thus consists of achieving the highest possible dividends with the lowest possible risk (low variance).

The total of all risk-adjusted dividends (in ECU) which you achieve via your portfolio decisions is multiplied by five at the end and then paid in euros.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way in order to achieve risk-adjusted dividend payments which are as high as possible.

Appendix A.2. Instructions (Treatment 'Representative')

You have the task of creating portfolios of shares. A portfolio of shares is a compilation of several shares.

The development of the share prices is of no concern to you, because you profit only once from the dividend payments of the shares in 2022. The dividend is the distribution of profits of a stock exchange-listed company to its shareholders.

You will receive information about how the dividend payments might turn out, and about the probabilities of different amounts of dividend. In addition, you will be shown how the dividends of the shares have developed over the last ten years.

You are paid the risk-adjusted dividend. A risk-adjusted dividend is the dividend payment divided by the variance of the dividend payments of the selected portfolio. Your task thus consists of achieving the highest possible dividends with the lowest possible risk (low variance).

The total of all risk-adjusted dividends (in ECU) which you achieve via your portfolio decisions is multiplied by five at the end and then paid in euros. However, this amount is not paid to you, but to another participant. If you make successful decisions, one of the other participants will have something to be pleased about. If you make unsuccessful decisions, one of the other participants will be annoyed.

At the same time, another participant is making the decisions which determine your payment. Who has made portfolio decisions for whom will be announced at the end of the session.

So please remember why you made which decisions. The other participant might want you to justify your decisions if the results are disappointing.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way in order to achieve risk-adjusted dividend payments which are as high as possible.

Appendix B. Test Questions

Appendix B.1. Test Questions (Treatment 'Self')

What is a share portfolio?

(a) A compilation of shares, bonds and derivative instruments.
(b) A compilation of shares. *(correct)*
(c) A compilation of various securities without shares.

What is a dividend?

(a) It is the opposite of a multiplication.
(b) It is a major military unit.
(c) It is the distribution of profits by a stock exchange-listed company to its shareholders. *(correct)*

What do you profit from?

(a) From increases in the price of the shares that I choose.
(b) From the risk-adjusted dividends of the shares that I choose. *(correct)*
(c) From increases in the price of the shares that I choose, and from the dividends.

How can the algorithm (robo advisor) be deployed?

(a) I have to use the robo advisor.
(b) The robo advisor is not available to me.
(c) I have a free choice between either making the portfolio decisions myself or delegating the task to a robo advisor which is specialised in this field. *(correct)*

Appendix B.2. Test Questions (Treatment 'Representative')

What is a share portfolio?

(a) A compilation of shares, bonds and derivative instruments.
(b) A compilation of shares. *(correct)*
(c) A compilation of various securities without shares.

From whose decisions do you profit?

(a) From my own decisions.
(b) From the decisions of all participants.
(c) From the decisions of the participant who makes the decisions for me. *(correct)*

What determines the payment of the person for whom you make the decisions?

(a) The changes in the prices of the shares that I choose.
(b) The risk-adjusted dividends of the shares that I choose. *(correct)*
(c) The increases in the price of the shares that I choose, and the dividends of the shares that I choose.

How can the algorithm (robo advisor) be deployed?

(a) I have to use the robo advisor.
(b) The robo advisor is not available to me.
(c) I have a free choice between either making the portfolio decisions myself or delegating the task to a robo advisor which is specialised in this field. *(correct)*

Appendix C. The Tasks

Appendix C.1. Task 1 (Treatment 'Self')

There are two shares to choose from: share Y and share Z. The dividend payments of the two companies are independent random processes with two possible configurations: ECU 8 and ECU 0, and with an expected value of ECU 4. In the table you can see how high the dividend payments of the two shares were in the last 10 years.

Table A1. Dividend payments of the shares in task 1 of treatment 'Self'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share Y	ECU 8	ECU 0	ECU 8	ECU 8	ECU 8	ECU 0	ECU 8	ECU 0	ECU 0	ECU 0	?
Share Z	ECU 0	ECU 0	ECU 8	ECU 0	ECU 8	ECU 8	ECU 0	ECU 0	ECU 8	ECU 8	?

You may choose two shares. As payment you receive the risk-adjusted dividends of the two selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selected, you thus receive the risk-adjusted dividends of 2 Y shares, of 2 Z shares, or of 1 Y share + 1 Z share. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (ECU 8 or ECU 0) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (ECU 8 or ECU 0) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

Now make your choice!

- o I will let the robo advisor decide;
  I will decide myself and choose:
- o 2 Y shares;
- o 2 Z shares;
- o 1 Y share + 1 Z share.

Appendix C.2. Task 2 (Treatment 'Self')

There are two shares to choose from: share X and share Q. The dividend payments of the two companies are independent random processes with two possible configurations: ECU 4 and ECU 0, and with an expected value of ECU 2. In the table you can see how high the dividend payments of the two shares were in the last 10 years.

Table A2. Dividend payments of the shares in task 2 of treatment 'Self'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share X	ECU 0	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 4	ECU 4	?
Share Q	ECU 0	ECU 4	ECU 4	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 4	ECU 0	?

You may choose four shares. As payment you receive the risk-adjusted dividends of the four selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selected, you thus receive the risk-adjusted dividends of 4 X shares, of 4 Q shares, of 3 X shares + 1 Q share, of 3 Q shares + 1 X share, or of 2 X shares + 2 Q shares. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (ECU 4 or ECU 0) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (ECU 4 or ECU 0) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

Now make your choice!

- o I will let the robo advisor decide;
  I will decide myself and choose:
- o 4 X shares;
- o 4 Q shares;
- o 3 X shares + 1 Q share;
- o 3 Q shares + 1 X share;
- o 2 Q shares + 2 X shares.

Appendix C.3. Task 3 (Treatment 'Self')

There are two shares from a specific sector of industry to choose from (share K and share L). In the table you can see how high the dividend payments of the two shares were in the last 10 years. When business is good in the sector, the dividend of share K is ECU

6, and that of share L is ECU 7. When business is poor in the sector, the dividend of share K is ECU 2, and that of share L is ECU 1. The business situation in the sector can vary from year to year and thus has to be viewed as a random process: the probability of the business situation being either good or poor in 2022 is 50% in each case.

Table A3. Dividend payments of the shares in task 3 of treatment 'Self'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share K	ECU 2	ECU 6	ECU 2	ECU 6	ECU 6	ECU 6	ECU 2	ECU 6	ECU 2	ECU 2	?
Share L	ECU 1	ECU 7	ECU 1	ECU 7	ECU 7	ECU 7	ECU 1	ECU 7	ECU 1	ECU 1	?

You may choose two shares. As payment you receive the risk-adjusted dividends of the two selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selected, you thus receive the risk-adjusted dividends of 2 K shares, of 2 L shares, or of 1 K share + 1 L share. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (good or poor economic situation in the sector) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (good or poor economic situation in the sector) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 2 K shares;
- o 2 L shares;
- o 1 K share + 1 L share.

Appendix C.4. Task 4 (Treatment 'Self')

There are two shares from a specific sector of industry to choose from (share M and share P). In the table you can see how high the dividend payments of the two shares were in the last 10 years. When business is good in the sector, the dividend of share M is ECU 4, and that of share P is ECU 3. When business is poor in the sector, the dividend of share M is ECU 0, and that of share P is ECU 1. The business situation in the sector can vary from year to year and thus has to be viewed as a random process: the probability of the business situation being either good or poor in 2022 is 50% in each case.

Table A4. Dividend payments of the shares in task 4 of treatment 'Self'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share M	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 0	ECU 4	?
Share P	ECU 3	ECU 1	ECU 3	ECU 1	ECU 1	ECU 1	ECU 3	ECU 3	ECU 1	ECU 3	?

You may choose four shares. As payment you receive the risk-adjusted dividends of the four selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selected, you thus receive the risk-adjusted dividends of 4 M shares, of 4 P shares, of 3 M shares + 1 P share, of 3 P shares + 1 M share, or of 2 M shares + 2 P shares. As the dividend payments are determined by a random process, it is not only the content

of the portfolio which determines your payment, but also luck. Which event (good or poor economic situation in the sector) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (good or poor economic situation in the sector) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 4 M shares;
- o 4 P shares;
- o 3 M shares + 1 P share;
- o 3 P shares + 1 M share;
- o 2 M shares + 2 P shares.

Appendix C.5. Task 1 (Treatment 'Representative')

There are two shares to choose from: share Y and share Z. The dividend payments of the two companies are independent random processes with two possible configurations: ECU 8 and ECU 0, and with an expected value of ECU 4. In the table you can see how high the dividend payments of the two shares were in the last 10 years.

Table A5. Dividend payments of the shares in task 1 of treatment 'Representative'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share Y	ECU 8	ECU 0	ECU 8	ECU 8	ECU 8	ECU 0	ECU 8	ECU 0	ECU 0	ECU 0	?
Share Z	ECU 0	ECU 0	ECU 8	ECU 0	ECU 8	ECU 8	ECU 0	ECU 0	ECU 8	ECU 8	?

You may choose two shares. As compensation, the risk-adjusted dividends are paid from the two selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selection, the risk-adjusted dividend of 2 Y shares, of 2 Z shares, or of 1 Y share + 1 Z share is paid out. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (ECU 8 or ECU 0) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (ECU 8 or ECU 0) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

The payment which you achieve with your decision is received by one of the other participants and not by you. This other participant might ask you to justify your choices, so you should think carefully about the decisions you make.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 2 Y shares;
- o 2 Z shares;
- o 1 Y share + 1 Z share.

Appendix C.6. Task 2 (Treatment 'Representative')

There are two shares to choose from: share X and share Q. The dividend payments of the two companies are independent random processes with two possible configurations: ECU 4 and ECU 0, and with an expected value of ECU 2. In the table you can see how high the dividend payments of the two shares were in the last 10 years.

Table A6. Dividend payments of the shares in task 2 of treatment 'Representative'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share X	ECU 0	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 4	ECU 4	?
Share Q	ECU 0	ECU 4	ECU 4	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 4	ECU 0	?

You may choose four shares. As compensation, the risk-adjusted dividends are paid from the four selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selection, the risk-adjusted dividend of 4 X shares, of 4 Q shares, of 3 X shares + 1 Q share, of 3 Q shares + 1 X share, or of 2 X shares + 2 Q shares is paid out. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (ECU 4 or ECU 0) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (ECU 4 or ECU 0) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

The payment which you achieve with your decision is received by one of the other participants and not by you. This other participant might ask you to justify your choices, so you should think carefully about the decisions you make.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 4 X shares;
- o 4 Q shares;
- o 3 X shares + 1 Q share;
- o 3 Q shares + 1 X share;
- o 2 Q shares + 2 X shares.

Appendix C.7. Task 3 (Treatment 'Representative')

There are two shares from a specific sector of industry to choose from (share K and share L). In the table you can see how high the dividend payments of the two shares were in the last 10 years. When business is good in the sector, the dividend of share K is ECU 6, and that of share L is ECU 7. When business is poor in the sector, the dividend of share K is ECU 2, and that of share L is ECU 1. The business situation in the sector can vary from year to year and thus has to be viewed as a random process: the probability of the business situation being either good or poor in 2022 is 50% in each case.

Table A7. Dividend payments of the shares in task 3 of treatment 'Representative'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share K	ECU 2	ECU 6	ECU 2	ECU 6	ECU 6	ECU 6	ECU 2	ECU 6	ECU 2	ECU 2	?
Share L	ECU 1	ECU 7	ECU 1	ECU 7	ECU 7	ECU 7	ECU 1	ECU 7	ECU 1	ECU 1	?

You may choose two shares. As compensation, the risk-adjusted dividends are paid from the two selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selection, the risk-adjusted dividend of 2 K shares, of 2 L shares, or of 1 K share + 1 L share is paid out. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (good or poor economic situation in the sector) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (good or poor economic situation in the sector) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

The payment which you achieve with your decision is received by one of the other participants and not by you. This other participant might ask you to justify your choices, so you should think carefully about the decisions you make.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 2 K shares;
- o 2 L shares;
- o 1 K share + 1 L share.

Appendix C.8. Task 4 (Treatment 'Representative')

There are two shares from a specific sector of industry to choose from (share M and share P). In the table you can see how high the dividend payments of the two shares were in the last 10 years. When business is good in the sector, the dividend of share M is ECU 4, and that of share P is ECU 3. When business is poor in the sector, the dividend of share M is ECU 0, and that of share P is ECU 1. The business situation in the sector can vary from year to year and thus has to be viewed as a random process: the probability of the business situation being either good or poor in 2022 is 50% in each case.

Table A8. Dividend payments of the shares in task 4 of treatment 'Representative'.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Share M	ECU 4	ECU 0	ECU 4	ECU 0	ECU 0	ECU 0	ECU 4	ECU 4	ECU 0	ECU 4	?
Share P	ECU 3	ECU 1	ECU 3	ECU 1	ECU 1	ECU 1	ECU 3	ECU 3	ECU 1	ECU 3	?

You may choose four shares. As compensation, the risk-adjusted dividends are paid from the four selected shares. The risk-adjusted dividend corresponds to the dividend payment divided by the variance of the dividend payments of the selected portfolio. Depending on the portfolio selection, the risk-adjusted dividend of 4 M shares, of 4 P shares, of 3 M shares + 1 P share, of 3 P shares + 1 M share, or of 2 M shares + 2 P shares is paid out. As the dividend payments are determined by a random process, it is not only the content of the portfolio which determines your payment, but also luck. Which event (good

or poor economic situation in the sector) occurs in the case of the two shares is determined separately by drawing lots for each round of the experimental survey.

You can make the portfolio decisions yourself or delegate them to an algorithm (robo advisor). The robo advisor is specialised in making meaningful portfolio decisions and takes all of the relevant information into account in an optimal way. However, the robo advisor also does not know which random event (good or poor economic situation in the sector) will occur as the dividend of the shares. In other words, even when the robo advisor is used, luck determines the payment to a certain extent.

The payment which you achieve with your decision is received by one of the other participants and not by you. This other participant might ask you to justify your choices, so you should think carefully about the decisions you make.

Now make your choice!

- o I will let the robo advisor decide;

  I will decide myself and choose:
- o 4 M shares;
- o 4 P shares;
- o 3 M shares + 1 P share;
- o 3 P shares + 1 M share;
- o 2 M shares + 2 P shares.

References

Agnew, Julie, Pierluigi Balduzzi, and Annika Sundén. 2003. Portfolio Choice and Trading in a Large 401(k) Plan. *The American Economic Review* 93: 193–215.

Alemanni, Barbara, Andrej Angelovski, Daniela Teresa di Cagno, Arianna Galliera, Nadia Linciano, Francesca Marazzi, and Paola Soccorso. 2020. Do Investors Rely on Robots? Evidence from an Experimental Study. *CONSOB Fintech Series* 7: 1–61.

Anderson, Anders. 2013. Trading and Under-Diversification. *Review of Finance* 17: 1699–741.

Andersson, Ola, Håkan J. Holm, Jean-Robert Tyran, and Erik R. Wengström. 2022. Deciding for Others Reduces Loss Aversion. *Management Science* 62: 29–36.

Back, Camila, Stefan Morana, and Martin Spann. 2021. *Do Robo-Advisors Make Us Better Investors?* Discussion Paper No. 276. München and Berlin: University of Munich (LMU) and Humboldt University Berlin, Collaborative Research Center Transregio 190—Rationality and Competition.

Baltussen, Guido, and Gerrit Thierry Post. 2011. Irrational Diversification: An Examination of Individual Portfolio Choice. *Journal of Financial and Quantitative Analysis* 46: 1463–91.

Barber, Brad M., and Terrance Odean. 2000. Trading is Hazardous to your Wealth: The Common Stock Investment Performance of Individual Investors. *Journal of Finance* 55: 773–806.

Beketov, Mikhail, Kevin Lehmann, and Manuel Wittke. 2018. Robo Advisors: Quantitative methods inside the robots. *Journal of Asset Management* 19: 363–70.

Benartzi, Shlomo 2001. Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock. *The Journal of Finance* 56: 1747–64.

Benartzi, Shlomo, and Richard H. Thaler. 2001. Naïve Diversification Strategies in Defined Contribution Saving Plans. *American Economic Review* 91: 79–98.

Bhatia, Ankita, Arti Chandani, and Jagriti Chhateja. 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: 1–9.

Blume, Marshall E., and Irwin Friend. 1975. The Asset Structure of Individual Portfolios and Some Implications for Utility Functions. *The Journal of Finance* 30: 585–603.

Bode, Matthias, Alexander van Echelpoel, and Christian R. Sievi. 1994. Multinationale Diversifikation: Viel zitiert, kaum befolgt. *Die Bank* 94: 202–6.

Bolton, Gary E., Axel Ockenfels, and Julia Stauf. 2015. Social responsibility promotes conservative risk behavior. *European Economic Review* 74: 109–27.

Boreiko, Dmitri, and Francesca Massarotti. 2020. How Risk Profiles of Investors Affect Robo-Advised Portfolios. *Frontiers in Artificial Intelligence* 3: 1–9.

Burns, Bruce D. 2001. The hot hand in basketball: Fallacy or adaptive thinking. *Proceedings of the Annual Meeting of the Cognitive Science Society* 23: 152–57.

Burton, Jason W., Mari-Klara Stein, and Tina Blegind Jensen. 2020. A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* 33: 220–39.

Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56: 809–25.

Chakravarty, Sujoy, Glenn W. Harrison, Ernan Haruvy, and E. Elisabet Rutstrom. 2011. Are You Risk Averse over Other People's Money? *Southern Economic Journal* 77: 901–13.

Charness, Gary, and Matthew. O. Jackson. 2009. The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization* 69: 241–47.

Chu, Zhong, Zhengwei Wang, Jing Jian Xiao, and Weiqiang Zhang. 2017. Financial literacy, portfolio choice and financial well-being. *Social Indicators Research* 132: 799–820.

Cornil, Yann, David J. Hardisty, and Yakov Bart. 2019. Easy, breezy, risky: Lay investors fail to diversify because correlated assets feel more fluent and less risky. *Organizational Behavior and Human Decision Processes* 153: 103–17.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G. Rossi. 2019. The Promises and Pitfalls of Robo-Advising. *The Review of Financial Studies* 32: 1983–2020.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144: 114–26.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2018. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 64: 1155–70.

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence. *Journal of Financial Economics* 119: 559–77.

Enke, Benjamin, and Florian Zimmermann. 2019. Correlation neglect in belief formation. *The Review of Economic Studies* 86: 313–32.

Eriksen, Kristoffer Wigestrand, Ola Kvaløy, and Miguel Luzuriaga. 2020. Risk-taking on behalf of others. *Journal of Behavioral and Experimental Finance* 26: 1–13.

Eyster, Erik, and Georg Weizsäcker. 2011. *Correlation Neglect in Financial Decision Making*. DIW Discussion Papers No. 1104. Berlin: German Institute for Economic Research (DIW).

Feier, Till, Jan Gogoll, and Matthias Uhl. 2022. Hiding behind machines: Artificial agents may help to evade punishment. *Science and Engineering Ethics* 28: 1–19.

Fernandes, Daniel. 2013. The 1/N Rule Revisited: Heterogeneity in the Naïve Diversification Bias. *International Journal of Research in Marketing* 30: 310–13.

Filiz, Ibrahim., Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2021a. Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance* 31: 1–8.

Filiz, Ibrahim, Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2021b. *The Tragedy of Algorithm Aversion*. WWP Wolfsburg Working Papers No. 21-02. Wolfsburg: Ostfalia University of Applied Sciences.

Fischbacher, Urs. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10: 171–78.

Füllbrunn, Sascha, and Wolfgang J. Luhan. 2015. *Am I My Peer's Keeper? Social Responsibility in Financial Decision Making*. Ruhr Economic Paper No. 551. Essen: RWI—Leibniz-Institut für Wirtschaftsforschung.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17: 295–314.

Goetzmann, William N., and Alok Kumar. 2008. Equity Portfolio Diversification. *Review of Finance* 12: 433–63.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai. 2021. Household finance. *Journal of Economic Literature* 59: 919–1000.

Gubaydullina, Zulia, and Markus Spiwoks. 2015. Correlation Neglect, Naïve Diversification, and Irrelevant Information as Stumbling Blocks for Optimal Diversification. *Journal of Finance and Investment Analysis* 4: 1–19.

Gubaydullina, Zulia, Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2021. *Creative Drive and Algorithm Aversion—The Impact of Influence in the Process of Algorithmic Decision-Making on Algorithm Aversion*. WWP Wolfsburg Working Papers No. 21-04. Wolfsburg: Ostfalia University of Applied Sciences.

Guiso, Luigi., Michael Haliassos, and Tullio Japelli. 2002. *Household Portfolios*. Cambridge: MIT Press.

Hedesstrom, Ted Martin, Henrik Svedsater, and Tommy Garling.. 2006. Covariation Neglect among Novice Investors. *Journal of Experimental Psychology: Applied* 12: 155–65.

Hibbert, Ann Marie, Edward R. Lawrence, and Arun J. Prakash. 2012. Can Diversification Be Learned? *The Journal of Behavioral Finance* 13: 38–50.

Hinsen, Silvana, Peter Hofmann, Jan Jöhnk, and Nils Urbach. 2022. How Can Organizations Design Purposeful Human-AI Interactions: A Practical Perspective from Existing Use Cases and Interviews. Paper presented at the 55th Hawaii International Conference on System Sciences (HICSS), Honolulu, HI, USA, January 4–7.

Hodge, Frank D., Kim I. Mendoza, and Roshan K. Sinha. 2021. The effect of humanizing robo-advisors on investor judgments. *Contemporary Accounting Research* 38: 770–92.

Huberman, Gur, and Paul Sengmueller. 2004. Performance and Employer Stock in 401(k) Plans. *Review of Finance* 8: 403–43.

Huberman, Gur, and Wei Jiang. 2006. Offering versus choice in 401(k) plans: Equity exposure and number of funds. *The Journal of Finance* 61: 763–801.

Jung, Dominik, Verena Dorner, Florian Glaser, and Stefan Morana. 2018. Robo-Advisory—Digitalization and Automation of Financial Advisory. *Business & Information Systems Engineering* 60: 81–86.

Jung, Markus, and Mischa Seiter. 2021. Towards a better understanding on mitigating algorithm aversion in forecasting: An experimental study. *Journal of Management Control* 32: 495–516.

Kallir, Ido, and Doron Sonsino. 2009. The Neglect of Correlation in Allocation Decisions. *Southern Economic Journal* 75: 1045–66.

Kawaguchi, Kohei. 2021. When will workers follow an algorithm? A field experiment with a retail business. *Management Science* 67: 1670–95.

Kim, Jungkeun., Marilyn Giroux, and Jacob C. Lee. 2021. When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology & Marketing* 38: 1140–55.

Lease, Ronald C., Wilbur G. Lewellen, and Gary G. Schlarbaum. 1974. The Individual Investor: Attributes and Attitudes. *The Journal of Finance* 29: 413–33.

Mahmud, Hasan, A.K.M. Najmul Islam, Syed Ishtiaque Ahmed, and Kari Smolander. 2022. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change* 175: 121390.

Méndez-Suárez, Mariano, Francisco García-Fernández, and Fernando Gallardo. 2019. Artificial intelligence modelling framework for financial automated advising in the copper market. *Journal of Open Innovation: Technology, Market, and Complexity* 5: 81.

Meulbroek, Lisa K. 2005. Company Stock in Pension Plans: How costly is it? *The Journal of Law and Economics* 48: 443–74.

Morrin, Maureen, J. Jeffrey Inman, Susan M. Broniarczyk, Gergana Y. Nenkov, and Jonathan Reuter. 2012. Investing for Retirement: The Moderating Effect of Fund Assortment Size on the 1/N Heuristic. *Journal of Marketing Research* 49: 537–50.

Niszczota, Paweł, and Dániel Kaszás. 2020. Robo-investment aversion. *PLoS ONE* 15: e0239277.

Pahlke, Julius, Sebastian Strasser, and Ferdinand M. Vieider. 2012. Risk-taking for others under accountability. *Economics Letters* 114: 102–5.

Pahlke, Julius, Sebastian Strasser, and Ferdinand M. Vieider. 2015. Responsibility effects in decision making under risk. *Journal of Risk and Uncertainty* 51: 125–46.

Polkovnichenko, Valery. 2005. Household Portfolio Diversification: A Case for Rank-dependent Preferences. *Review of Financial Studies* 18: 1467–502.

Polman, Evan. 2012. Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes* 119: 141–50.

Prahl, Andrew, and Lyn Van Swol. 2017. Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting* 36: 691–702.

Reynolds, Douglas B., Jacob Joseph, and Reuben Sherwood. 2009. Risky Shift Versus Cautious Shift: Determining Differences in Risk Taking between Private and Public Management Decision-Making. *International Journal of Economics and Business Research* 7: 63–78.

Roberts, Harry V. 1959. Stock market "patterns" and financial analysis: Methodological suggestions. *Journal of Finance* 1: 1–10.

Robillard, Joseph. 2018. Millennial Attitudes towards Financial Advisors and Emerging Investment Technologies. *Wharton Research Scholars* 171: 1–40. Available online: https://repository.upenn.edu/wharton_research_scholars/171 (accessed on 1 July 2022).

Rogers, Paul. 1998. The cognitive psychology of lottery gambling: A theoretical review. *Journal of Gambling Studies* 14: 111–34.

Rossi, Alberto G., and Stephen P. Utkus. 2020. Who Benefits from Robo-Advising? Evidence from Machine Learning. SSRN Working Paper. Available online: https://ssrn.com/abstract=3552671 (accessed on 1 July 2022).

Rühr, Alexander. 2020. Robo-Advisor Configuration: An Investigation of User Preferences and the Performance-Control Dilemma. Paper presented at the 28th European Conference on Information Systems (ECIS), Online, June 15–17.

Rühr, Alexander, Benedikt Berger, and Thomas Hess. 2019a. Can I Control My Robo-Advisor? Trade-Offs in Automation and User Control in (Digital) Investment Management. Paper presented at the 25th Americas Conference on Information Systems (AMCIS), Cancún, Mexico, August 15–17, pp. 1–10.

Rühr, Alexander, Benedikt Berger, and Thomas Hess. 2019b. A Classification of Decision Automation and Delegation in Digital Investment Systems. Paper presented at the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, January 8–11, pp. 1435–44.

Singh, Ishmeet, and Navjot. Kaur. 2017. Wealth Management through Robo Advisory. *International Journal of Research—Granthaalayah* 5: 33–43.

Taylor, Earl L. 2017. Making sense of "algorithm aversion". *Research World* 64: 57–57.

Tversky, Amos, and Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76: 105–10.

Uhl, Matthias W., and Philippe Rohner. 2018. Robo-advisors versus traditional investment advisors: An unequal game. *The Journal of Wealth Management* 21: 44–50.

Ungeheuer, Michael, and Martin Weber. 2021. The perception of dependence, investment decisions, and stock prices. *The Journal of Finance* 76: 797–844.

Vieider, Ferdinand, Clara Villegas-Palacio, Peter Martinsson, and Milagros Mejía. 2016. Risk taking for oneself and others: A structural model approach. *Economic Inquiry* 54: 879–94.

von Walter, Benjamin, Dietmar Kremmel, and Bruno Jäger. 2021. The impact of lay beliefs about AI on adoption of algorithmic advice. *Marketing Letters* 33: 143–55.

Wärneryd, Karl-Erik 2001. *Stock-Market Psychology*. Cheltenham: Edward Elgar.

Zielonka, Piotr. 2004. Technical analysis as the representation of typical cognitive biases. *International Review of Financial Analysis* 13: 217–25.

Chapter VII

Interest Rate Forecasts in Latin America

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Published in:
Journal of Economic Studies, 49(5), 920-936.

Also published in:
Sofia Diskussionsbeiträge, Sonderforschungsgruppe Institutionenanalyse, No. 20-5, Darmstadt, September 2020. (Discussion Paper in German)
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 20-02, Wolfsburg, October 2020. (Working Paper)

Abstract

Purpose – This paper aims to assess the quality of interest rate forecasts for the money markets in Argentina, Brazil, Chile, Mexico and Venezuela for the period between 2001 and 2019. Future interest rate trends are of key significance for many business-related decisions. Thus, reliable interest rate forecasts are essential, for example, for banks that make profits by carrying out maturity transformations.

Design/methodology/approach – The data that we analyze were collected by Consensus Economics through a monthly survey with over 120 renowned economists and were published between 2001 and 2019 in the journal Latin American Consensus Forecasts. The authors use the Diebold-Mariano test, the sign accuracy test, the TOTA coefficient and the unbiasedness test to determine the precision and biasedness of the forecasts.

Findings – The research reveals that the forecasting work carried out in Brazil, Chile and Mexico is remarkably successful. The quality of forecasts from Argentina and Venezuela, on the other hand, is significantly poorer.

Originality/value – Over 50 studies have already been published with regard to the accuracy of interest rate forecasts, emphasizing the importance of the topic. However, interest rate forecasts for Latin American money markets have hardly been considered thus far. The paper closes this research gap. Overall, the analyzed database amounts to a total of 209 forecast time series with 28,451 individual interest rate forecasts. This study is thus far more comprehensive than all previous studies.

Keywords

Forecast accuracy, interest rate forecasts, maturity transformation.

JEL Classification

E44, E47, F37, G15, G17, G21.

Chapter VIII

Sticky Stock Market Analysts

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Published in:
Journal of Risk and Financial Management, 14(12), 593.

Also published in:
Sofia Diskussionsbeiträge, Sonderforschungsgruppe Institutionenanalyse, No. 21-3, Darmstadt, February 2021. (Discussion Paper in German)
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 21-03, Wolfsburg, April 2021. (Working Paper)

Abstract

Technological progress in recent years has made new methods available for making forecasts in a variety of areas. We examine the success of ex-ante stock market forecasts of three major stock market indices, i.e., the German Stock Market Index (DAX), the Dow Jones Industrial Index (DJI), and the Euro Stoxx 50 (SX5E). We test whether the forecasts prove true when they reach their effective dates and are therefore suitable for active investment strategies. We revive the thoughts of the American sociologist William Fielding Ogburn, who argues that forecasters consistently underestimate the variability of the future. In addition, we draw on some contemporary measures of forecast quality (prediction-realization diagram, test of unbiasedness, and Diebold–Mariano test). We reveal that (a) unusual events are underrepresented in the forecasts, (b) the dispersion of the forecasts lags behind that of the actual events, (c) the slope of the regression lines in the prediction-realization diagram is <1, (d) the forecasts are highly biased, and (e) the quality of the forecasts is not significantly better than that of naïve forecasts. The overall behavior of the forecasters can be described as "sticky" because their forecasts adhere too strongly to long-term trends in the indices and are thus characterized by conservatism.

Keywords

Stock market forecasting, forecasting bias, variability of reality, conservatism of predictors.

JEL Classification

D83, D84, D91, G17, G41.

Journal of
Risk and Financial Management

MDPI

Article

Sticky Stock Market Analysts

Ibrahim Filiz [1,*], Jan René Judek [1], Marco Lorenz [2] and Markus Spiwoks [1]

[1] Faculty of Business, Ostfalia University of Applied Sciences, Siegfried-Ehlers-Street 1, D-38440 Wolfsburg, Germany; ja.judek@ostfalia.de (J.R.J.); m.spiwoks@ostfalia.de (M.S.)
[2] Faculty of Economic Sciences, Georg August University Göttingen, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany; marco.lorenz@stud.uni-goettingen.de
* Correspondence: ibrahim.filiz@ostfalia.de; Tel.: +49-160-3344-078

Abstract: Technological progress in recent years has made new methods available for making forecasts in a variety of areas. We examine the success of ex-ante stock market forecasts of three major stock market indices, i.e., the German Stock Market Index (DAX), the Dow Jones Industrial Index (DJI), and the Euro Stoxx 50 (SX5E). We test whether the forecasts prove true when they reach their effective dates and are therefore suitable for active investment strategies. We revive the thoughts of the American sociologist William Fielding Ogburn, who argues that forecasters consistently underestimate the variability of the future. In addition, we draw on some contemporary measures of forecast quality (prediction-realization diagram, test of unbiasedness, and Diebold–Mariano test). We reveal that (a) unusual events are underrepresented in the forecasts, (b) the dispersion of the forecasts lags behind that of the actual events, (c) the slope of the regression lines in the prediction-realization diagram is <1, (d) the forecasts are highly biased, and (e) the quality of the forecasts is not significantly better than that of naïve forecasts. The overall behavior of the forecasters can be described as "sticky" because their forecasts adhere too strongly to long-term trends in the indices and are thus characterized by conservatism.

Keywords: stock market forecasting; forecasting bias; variability of reality; conservatism of predictors

JEL Classification: D83; D84; D91; G17; G41

Citation: Filiz, Ibrahim, Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2021. Sticky Stock Market Analysts. *Journal of Risk and Financial Management* 14: 593. https://doi.org/10.3390/jrfm14120593

Academic Editor: Goran Buturac

Received: 10 November 2021
Accepted: 3 December 2021
Published: 9 December 2021

1. Introduction

Capital market forecasts often show a closer connection to the capital market development of the present than to the capital market development of the future. This phenomenon is known as topically orientated trend adjustment (Andres and Spiwoks 1999). It occurs equally in share price forecasts, interest rate forecasts, exchange rate forecasts, and commodity price forecasts (see, e.g., Filiz et al. 2019; Kunze et al. 2018; Spiwoks et al. 2015; Spiwoks and Hein 2007). A tendency to underestimate the variability of reality could be an important cause (Spiwoks et al. 2015).

The American sociologist William Fielding Ogburn discovers almost 90 years ago that forecasters systematically underestimate the actual variability of reality (Ogburn 1934). He provides a concrete research approach to identify such behavior. Presumably because Ogburn deals with the prognosis of sporting events and not with the prognosis of economic events, he has so far not been noticed by economic research.

During an empirical analysis of the forecasting behavior of experts and lay people, Ogburn (1934) concludes that the variability of reality is consistently underestimated. He traces this back to a tendency which he calls the "conservatism of the predictors". In detail, he is referring to:

1. Unusual events (e.g., a sudden drop in an otherwise rising trendline) are forecasted more seldom than they occur in reality, whereas normal events (e.g., a recently rising trendline continuing to rise) are over-represented in forecasts.

2. The standard deviation of the forecasts is lower than the standard deviation of the actual events.
3. The extent of the forecasted changes lags behind the scale of the actual changes.

Active investment strategies have been popular since the emergence of modern stock markets (Maxwell and van Vuuren 2019; Lofthouse 1996; Friend and Vickers 1965; Cowles 1933). In order to successfully design active investment strategies such as market timing, stock picking, or index picking, forecasts of future stock market developments are indispensable. New forecasting methods are constantly being discussed: econometric models (Goyal et al. 2021; Chen and Vincent 2016; Welch and Goyal 2008), artificial neural networks (Rajab and Sharma 2019; Atsalakis and Valavanis 2009), artificial intelligence (Mallikarjuna and Rao 2019), capital market simulations with multi-agent models (Yang et al. 2020; Krichene and El-Aroui 2018; Arthur et al. 1997), modelling based on the expectations of capital market agents (Atmaz et al. 2021; Greenwood and Shleifer 2014), and neuro-psycho-economics approaches (Ortiz-Teran et al. 2019; Kandasamy et al. 2016; Werner et al. 2009). However, testing these approaches using ex-post forecasts in an out-of-sample data domain repeatedly leads to apparent forecasting successes that then may not materialize in real ex-ante settings (Kazak and Pohlmeier 2019). When the variability of reality is systematically underestimated, this can contribute towards very costly errors in the field of stock market forecasts. Under certain circumstances, basing active investment strategies on inappropriate stock market forecasts can lead to serious losses and even bankruptcy, when expected returns do not occur. Due to the necessity of reliable forecasts for a successful active investment strategy, stock market forecasting is a dynamic field of research.

The reliability of stock market forecasts is rarely examined. There are many studies on pre-tax profit forecasts (Ramnath et al. 2008), but research on the success of actual ex-ante forecasts in stock prices, stock market indices, or stock market returns are still a rarity. So far, it has not been in the focus of research whether stock market forecasts are characterized by a systematic underestimation of the variability of reality as found by Ogburn (1934). This research gap is even more surprising because the necessary investigation tools have long been available in the form of Theil's prediction-realization diagram and the test for unbiasedness. We raise the question of how successful experts were in forecasting major stock indices (DAX, Dow Jones Industrial Index, Euro Stoxx 50) in the period from 1992 to 2020. We use Ogburn's (1934) examination instruments. But we also go beyond this and use current standard procedures such as the comparison to the naïve forecasts (Diebold–Mariano test) and the unbiasedness test.

The forecasts turn out to be quite unreliable. Indeed, forecasters underestimate the variability of reality. This offers interesting starting points for improving the forecasting process.

2. Literature Review

2.1. Technological Progress in Stock Market Forecasting

There is a rich literature on the appliance of advanced econometric methodology in the forecasting process in order to identify meaningful predictors for future events. Guo (2006) uses ordinary and dynamic least squares regressions to analyze whether four different variables can be used as predictors for stock returns. The study concludes that the consumption-wealth ratio can indeed be used for statistically significant forecasts. Chen and Vincent (2016) also use different econometric models applied to full-sample approaches and out-of-sample approaches in order to analyze the informational value of different variables for the development of the Standard and Poor's 500 index (S&P 500) for the period 1964 to 2011. They conclude that the market momentum and the investor sentiment can indeed serve as potential predictors for bear markets. In a similar study, Neely et al. (2014) find that adding technical variables to the commonly used macroeconomic predictors can significantly improve the quality of forecasts for the equity risk premium.

Welch and Goyal (2008) examine the informative value of 13 frequently used variables such as dividend yields or inflation. In contrast to the researchers mentioned above, they find that none of the 13 variables can be used to predict the S&P 500 index returns from 1926 to 2004 neither in-sample nor out-of-sample. Quite importantly, they also find that none of the information available at the time of a potential investment decision would have helped to gain an idea of future developments. A couple of years later, the same authors extend their research to 29 additional variables that have been brought up in the discussion in the meantime. In spite of the advances in research methods, they still diagnose a poor usefulness in predicting the equity premium in-sample and out-of-sample (Goyal et al. 2021).

Bahrami et al. (2018) add to the research by finding that even though most variables themselves do not lead to significant forecasts, combining forecasts from individual predictive models significantly improves the quality of stock return forecasts for ten advanced emerging markets across the globe.

Whereas most studies cited above apply OLS regression models, Nyberg (2013) examines the suitability of dynamic binary time series models for predicting the S&P 500 index between 1957 and 2010. The author concludes that both in-sample and out-of-sample, dynamic binary time series models are able to successfully forecast bull and bear markets.

A very dynamic research area is capital market simulation with multi-agent models. Heterogeneous agents interact with one another on an artificial stock market. Their demand for shares and their supply of shares are brought together in a stock exchange, so that the development of the share prices results from the actions of the individual agents. These in turn observe the development of the share price and adjust their further behavior to the development of the share price. In this way, the special dynamics of interactions on stock markets can be modeled and examined more closely. The artificial stock markets are validated using the stylized facts (e.g., fat tails, gain-loss asymmetry, volatility clustering, volume-volatility correlation). The price patterns of artificial stock markets should correspond to the price patterns of real stock markets.

The first highlight of this research area is the Santa Fe Artificial Stock Market (Arthur et al. 1997). The Frankfurt Artificial Stock Market (Hein et al. 2012) also takes into account a realistic stock exchange mechanism, different communication structures between the agents, and different investment philosophies of the agents. Recently, for example, information asymmetries (Krichene and El-Aroui 2018), memory length and confidence level (Bertella et al. 2014), risk preference (Chen and Huang 2008), tick size systems (Yang et al. 2020), and different types of stocks (Ponta and Cincotti 2018) have been taken into account in artificial stock markets. Artificial stock markets have the significant advantage that extreme events (crashes) can be observed more frequently and can be better analyzed than on real stock markets. The decisive disadvantage of the artificial stock markets is that the models are still too abstract to lead to very concrete share price forecasts.

Another very dynamic research area uses survey data to examine the expectations of capital market players more closely (e.g., Atmaz et al. 2021; Cassella and Gulen 2019; Cassella and Gulen 2018; Greenwood and Shleifer 2014). In some approaches, different types of investors (lay people vs. professionals or contrarians vs. extrapolators) are taken into account. The different expectations of these investor groups are then used to develop models for describing or forecasting share price developments. These approaches appear particularly promising because the special importance of the expectations for capital market events is emphasized. In addition, real capital market data are linked with survey data on the expectations of capital market players in a very differentiated manner. In contrast to the approaches of capital market simulation based on multi-agent models, these research approaches remain close to the observable reality of price formation on the stock markets.

In recent years, there have also been promising results regarding neuro-fuzzy systems used for stock price forecasting. For example, Atsalakis and Valavanis (2009) create a neuro-fuzzy system that outperforms a traditional "buy and hold"-strategy regarding the Athens and the New York Stock Exchange. Even in a direct comparison to econometric methods, Rajab and Sharma (2019) show that neuro-fuzzy approaches to forecasting the Bombay

Stock Exchange, CNX Nifty, and S&P 500 can significantly outperform multiple regression analysis models or generalized autoregressive conditional heteroscedasticity models.

On the other hand, Mallikarjuna and Rao (2019) find that traditional linear and non-linear models are more accurate at forecasting daily stock market returns of selected indices from developed, emerging, and frontier markets for the period 2000 to 2018 than newly emerged artificial intelligence and frequency domain models. However, neither of the four models nor hybrid approaches provide satisfying results across the markets in their study.

In the field of neuro-psycho-economic approaches, Kandasamy et al. (2016) show that interoception, i.e., the perception of physiological signals from within the body, seems to play a role in the success of professional financial traders. Werner et al. (2009) also show that people with good cardiac perception perform better when choosing between profit and loss options.

In the context of ex-post forecasts in the out-of-sample area, these approaches sometimes show enormous potential. However, many of these approaches have yet to prove their suitability for actual ex-ante forecasts. Their informative value for ex-ante forecasts might be limited due to, for example, differences in estimation risk and low statistical power (Kazak and Pohlmeier 2019).

2.2. Ex-Ante Stock Market Forecasts

The actual success of stock market forecasts is thus best checked against real ex-ante forecasts. In the area of interest rate forecasts, the evaluation of continuously published forecasts has a long tradition (Filiz et al. 2021; Fassas et al. 2021; Filiz et al. 2019; Kunze et al. 2017; Miah et al. 2016; Pierdzioch 2015; Baghestani et al. 2015; Oliver and Pasaogullari 2015; Spiwoks et al. 2015). In the area of stock market forecasting, however, there are only a small number of studies that check continuously published stock market forecasts for their reliability (see the synoptic overview in Table 1).

Lakonishok (1980) analyzes forecasts for the S&P 425 index in the period from 1947 to 1974. He concludes that the reliability of the forecasts does not go recognizably beyond that of naïve forecasts. In this context, a naïve forecast is defined as the assumption that the prevailing value for the variable being forecast at the time the forecast is made will also prevail in the future. In addition, the forecasts are biased and systematically underestimate the returns of the S&P 425. Dimson and Marsh (1984) analyze the forecasted returns of 206 selected British shares in the period from 1980 to 1981. The authors conclude that the forecasts are successful and can lead to systematic excess returns. Fraser and MacDonald (1993) examine forecasts for the development of the French CAC 40 index in the period from 1984 to 1987. This reveals that the forecasts are less reliable than naïve forecasts. Furthermore, it is evident that the forecasts tend to be oriented towards the present rather than the future.

Spiwoks (2004) and Spiwoks and Hein (2007) consider forecasts for six international share indices (the Dow Jones Industrial Index, the DAX, the FT-SE 100, the CAC 40, MIBtel, and the Nikkei 225) issued in the period from 1994 to 2004. The results are very similar. Almost without exception, the forecast time series exhibit greater forecasting errors than the respective naïve forecast. In addition, they exhibit topically orientated trend adjustment (Andres and Spiwoks 1999). In other words, they reflect the present situation more than anything else, and hardly provide any insights into future trends.

Benke (2006) examines DAX forecasts for the period from 1992 to 2005. He establishes that the forecasters consistently underestimate the extent of the actual changes. Bacchetta et al. (2009) analyze forecasts for the Dow Jones Industrial Index and the Nikkei 225 in the period from 1998 to 2005. The authors conclude that the forecasts are suitable for achieving systematic excess returns. Fujiwara et al. (2013) observe TOPIX forecasts in the years from 1998 to 2010. They argue that the forecasters are too strongly orientated towards their previous forecasts and systematically underestimate the actual trends of the TOPIX.

Table 1. Synoptic overview of studies on ex-ante stock market forecasts.

Study	Subject of the Forecast	Methods	Time Scale	Result
Lakonishok (1980)	S&P 425	Unbiasedness test with Theil–Sen estimator, Theil's U, turning point errors	1947–1974	–
Dimson and Marsh (1984)	Selected British shares	Comparison of forecast and actual return via *t*-test, Unbiasedness test	1980–1981	+
Fraser and MacDonald (1993)	CAC 40	Unbiasedness test, root mean squared error	1984–1987	–
Spiwoks (2004)	Dow Jones Industrial Index, DAX, FT-SE 100, CAC 40, MIBtel, and the Nikkei 225	Analysis of turning point errors, Theil's U, TOTA coefficient	1994–2004	–
Benke (2006)	DAX	Comparison of absolute frequencies regarding forecasting errors, direction of error, and comparison to naïve forecasts without statistical test	1992–2005	–
Spiwoks and Hein (2007)	Dow Jones Industrial Index, DAX, FT-SE 100, CAC 40, MIBtel, and the Nikkei 225	Root mean squared relative error, mean absolute relative error	1994–2004	–
Bacchetta et al. (2009)	Dow Jones Industrial Index, and Nikkei 225	Log Regression	1998–2005	+
Fujiwara et al. (2013)	TOPIX	Augmented Dickey–Fuller test, ADF-Fisher chi-square test	1998–2010	–

+ = Overall, the forecasts are assessed as good; – = overall, the forecasts are assessed as being flawed.

As we want to consider the abilities of professional stock market analysts, experimental studies in which the subjects are asked to make stock market forecasts themselves (e.g., Theissen 2007; De Bondt 1993) are not considered here.

2.3. Hypotheses

Capital market forecasts often describe the present rather than the future. Spiwoks et al. (2015) cite the systematic underestimation of the variability of reality as a possible reason for the phenomenon of topically oriented trend adjustments in capital market forecasts. The American sociologist William Fielding Ogburn (1934) is the first to address the systematic underestimation of the variability of reality in predicting future events. He presumes that (1) unusual events (e.g., a sudden drop in an otherwise rising trendline) are forecasted too seldom, that (2) the standard deviation of the forecasts is lower than the standard deviation of the actual events, and that (3) the forecasted changes lag behind the actual changes.

We check whether the forecasts for the German Stock Market Index (DAX), the Dow Jones Industrial Index (DJI) and the Euro Stoxx 50 (SX5E) also show these three properties. In formulating the hypotheses, we assume that the observations made by Ogburn (1934) who investigated forecasts of sporting events also apply to stock market forecasts.

Unlike the DAX, the DJI and the SX5E are price indices. Nevertheless, their long-term development is considered to be non-stationary. Over the long term, a rising trend can be recognized in all three stock indices. To this extent, it is simple to define unusual and normal events. A normal event is an increase in the share price index. An unusual event is a decrease in the share price index. Hypotheses 1 and 2 are therefore:

Hypothesis 1. *Falls in stock market indices are forecasted more seldom than they occur in reality.*

Hypothesis 2. *The standard deviation of the forecasted changes of the stock market indices is lower than the standard deviation of the actual changes in the indices.*

Should the systematical underestimation of the variability of reality be true in our data basis, investors would be exposed to a high risk, as relatively large changes in trends, also negative ones, would not be reflected adequately in the forecasts. The best way to test this assumption is to compute a prediction-realization diagram (Theil 1958) that compares the forecasted relative share price changes to the actual relative share price changes (as described in the Methods section). If the forecast changes are smaller than the actual changes, this leads to a regression line with a slope of <1 in the prediction-realization diagram. Hypothesis 3 therefore reads:

Hypothesis 3. *The slope of the regression lines in the prediction-realization diagram is lower than one (slope < 1).*

If the predicted changes lag behind the actual changes and it is thus true that the forecasters are guided by conservatism, the forecasts are not unbiased. This can be verified best by means of the test of unbiasedness using the Mincer–Zarnowitz regression (as described in the Methods section). The use of the unbiasedness test is of particular interest here because it can be used to determine whether the underestimation of the changes in the prognosis object can be viewed as statistically significant. Hypothesis 4 is therefore:

Hypothesis 4. *The forecasts prove to be biased.*

An assessment of capital market forecasts is incomplete if the forecasts are not compared to the naïve forecasts. In view of the results of previous studies (Spiwoks and Hein 2007; Spiwoks 2004; Fraser and MacDonald 1993; Lakonishok 1980), we expect that the quality of the forecasts will not be significantly better than that of naïve forecasts. If this is the case, investors should by no means consider the forecasts, as the naïve forecast is readily available at any time. Hypothesis 5 is therefore:

Hypothesis 5. *The quality of the forecasts is not significantly higher than that of naïve forecasts.*

3. Data Basis

We evaluate DAX forecasts which were published between 1992 and 2020 in the *Handelsblatt* newspaper (HB). The forecasts have a forecast horizon of one year. In addition, we evaluate forecasts for the DAX and the Euro Stoxx 50 which were published in the period from 2002 to 2020 in the *Frankfurter Allgemeine Zeitung* (FAZ). We also analyze forecasts for the Dow Jones Industrial Index which were published between 2004 and 2020 in the FAZ. The time scales differ as we have taken into account all stock price forecasts since the beginning of their publication in order to get more meaningful results. These forecasts have forecast horizons of six and twelve months (Table 2). We provide the dataset used in our study as a Supplementary in an Excel format. The dataset comprises all analyzed forecasts published annually in the *Frankfurter Allgemeine Zeitung* and *Handelsblatt* between 1992 and 2020.

Table 2. Data basis and summary statistics.

Source	Subject	Period	N	Min (in %)	Max (in %)	Median (in %)	Mean (in %)	N	Min (in %)	Max (in %)	Median (in %)	Mean (in %)
			Forecast Horizon 6 Months					Forecast Horizon 12 Months				
HB	DAX	1992–2020	NA	NA	NA	NA	NA	964	−25.16	72.85	8.08	8.76
FAZ	DAX	2002–2020	282	−33.47	18.68	3.38	2.34	402	−25.16	45.20	8.14	8.94
FAZ	DJI	2004–2020	203	−21.45	23.06	1.62	1.39	259	−20.24	42.43	6.07	5.95
FAZ	SX5E	2002–2020	270	−34.63	22.57	3.24	2.32	381	−20.33	36.87	7.88	8.03
Σ			755					2006				

HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; DAX = German Stock Market Index; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; N = number of forecasts issued; Min = minimum; Max = maximum; NA = not available.

In Table 2, we also provide descriptive statistics and show both the minimum and maximum predicted percentage index level changes as well as the median and mean value of the predicted percentage index level changes for the examined data. The descriptive statistics on forecast index level changes in Table 2 are shown in percentages to give a clearer picture of the data. The institutes did not forecast percentage index level changes, but rather the respective index levels. For example, M.M. Warburg & Co. predicted the DAX index level at the end of the year 2009 at 3600 points. At the time the forecast was issued, the DAX had an index level of 4810.20 points. Thus, the institute forecast the largest price decline of 25.16%. The WGZ-Bank forecast the maximum percentage increase in the index level of the DAX in 2003. While the DAX had an index level of 2892.63 points at the time the forecast was made, the bank forecast a percentage increase of 72.85% to 5000 points at the end of the year. On average, the institutes forecast an index level increase of the DAX of 8.76% (median 8.08%) in the period considered from 1992 to 2020 (see Table A1 in Appendix A for more detailed descriptive statistics on our data basis). In Figure 1, we provide an overview of the 12-month forecasts examined by showing the mean values of the forecasts, the associated actual index values, and the naïve forecasts.

The forecasts are from private German banks such as Fürst Fugger Privatbank or Bethmann Bank, from German state banks such as Helaba or Bayerische Landesbank, from major German banks such as Deutsche Bank or Commerzbank, and from international banks like Goldman Sachs, J.P. Morgan, or BNP Paribas. For a detailed overview of which institutes published forecasts in which newspaper, see Appendices B and C.

The methods applied by the individual institutions in order to obtain their forecasts are not disclosed. The forecasts are collected by HB and FAZ through annual quantitative surveys. For example, at the end of 2019, the newspapers collected and published forecasts that were drawn up for the middle and the end of 2020.

To the best of our knowledge, an analysis of the quality of actual ex-ante forecasts for the Euro Stoxx 50 has not yet been the subject of the literature (Table 1). Ex-ante forecasts of the Dow Jones Industrial Index and the DAX have also not been considered since 2005. Since then, technological progress has led to the emergence of numerous new forecasting tools and methods, which are discussed in our literature section. Overall, our data basis consists of 2761 forecasts covering a period of time of up to 28 years per time series. We are therefore convinced that an analysis of this data basis is a useful addition to the existing literature on stock market forecasts.

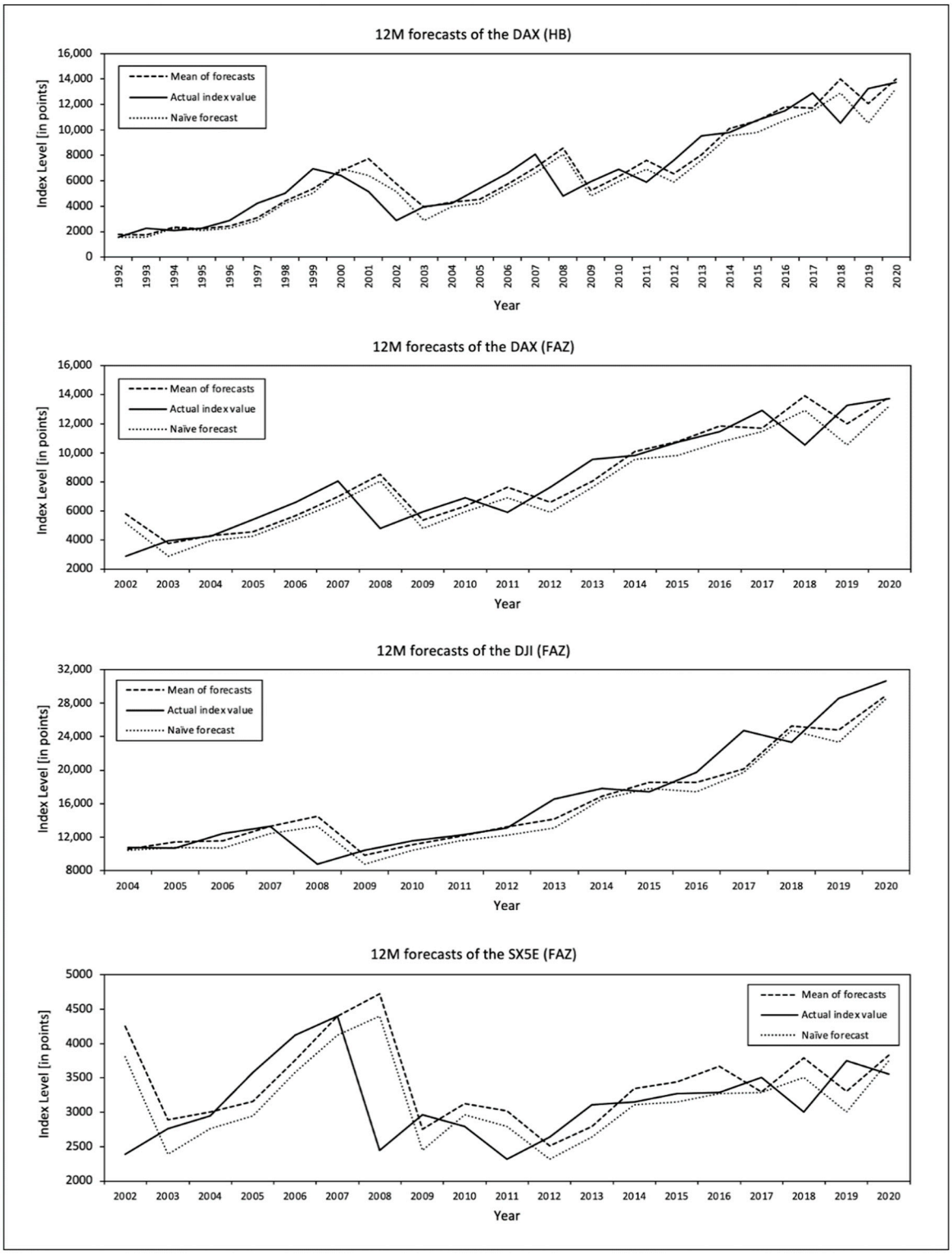


Figure 1. Means of 12M forecasts, actual index values, and naïve forecasts of the DAX, DJI, and SX5E.

4. Methods

Fundamentally, we follow Ogburn's assessment of forecasting: Ogburn (1934) assumes that forecasters suffer from conservatism. Therefore, we examine whether (1) unusual events are forecast too infrequently, (2) the standard deviation of the forecasts is lower than the standard deviation of the actual events, and (3) forecast changes lag behind actual

changes. We consider these three aspects in the forecasts as a whole, but also individually for all forecasters who issue forecasts for at least ten years. In addition, we also go beyond Ogburn's methodology and include some contemporary additions to address the assessment of forecast quality from today's perspective. As statistical tools to measure the quality of the survey-based forecasts we use Theil's prediction-realization diagram (Theil 1958), the test for the unbiasedness of the forecasts, and the Diebold–Mariano test for a comparison to the respective naïve forecast.

We draw on the prediction-realization diagram for a qualitative assessment of forecasting errors. For this purpose, we first calculate the forecasted relative changes (ρPF) and the realized relative actual stock price changes (ρPA). A_t shows the actual event at the time for which the forecast is applied and A_{t-h} shows the actual event at the time when the forecast was made.

$$\rho PF = \frac{P_t - A_{t-h}}{A_{t-h}} \text{ and } \rho PA = \frac{A_t - A_{t-h}}{A_{t-h}} \tag{1}$$

P = forecast of the actual event;
A = actual event;
t = time;
h = forecast horizon.

The forecasted percentage changes and the actual percentage changes are plotted and compared in the prediction-realization diagram (Figure 2). The dashed diagonal line in the prediction-realization diagram reflects the area in which the forecasted percentage changes and the actual realized percentage changes coincide (perfect forecasts). A good forecast time series is therefore characterized by the fact that the values are close to the diagonal. Using an OLS regression, we examine whether the slope of the regression line resulting from the forecasts considered is equal to one. When the variability of actual events is systematically underestimated, the slope of the regression lines in the prediction-realization diagram should be lower than one. A flat course of the regression lines (slope < 1) indicates an underestimation of the actual changes.

For all forecasters who have been taking part in forecasting surveys for at least ten years, we determine the slope of the regression lines individually. All of the other forecasts are evaluated within the framework of the total number of forecasts analyzed and within the framework of the consensus forecasts.

Furthermore, we perform the unbiasedness test using the Mincer–Zarnowitz regression (Mincer and Zarnowitz 1969) to examine whether forecasting errors are systematic. The Mincer–Zarnowitz regression takes the following form:

$$A_t = \alpha + \beta P_t + u_t \tag{2}$$

A_t = event that actually occurred in time t (dependent variable);
α = constant;
β = coefficient of the respective forecast;
P_t = forecast of the actual event in time t;
u_t = error term in time t.

Based on this equation, forecasts are considered unbiased if α is not significantly different to 0, and β is not significantly different to 1. Likewise, the error term u_t may not be autocorrelated. Forecasts are considered unbiased when, with a low probability of error, the joint hypothesis of $\alpha = 0$ and $\beta = 1$ does not have to be rejected. This is checked by using the Wald test (Wald 1943). A further condition is the absence of autocorrelation in the values of the error term u_t, which is examined with the Wooldridge test (Wooldridge 2002). If, according to these criteria, a forecast time series is unbiased, Granger and Newbold (1974) argue that this by no means signifies that the forecasts are perfect. They merely do not exhibit any *systematic* errors.

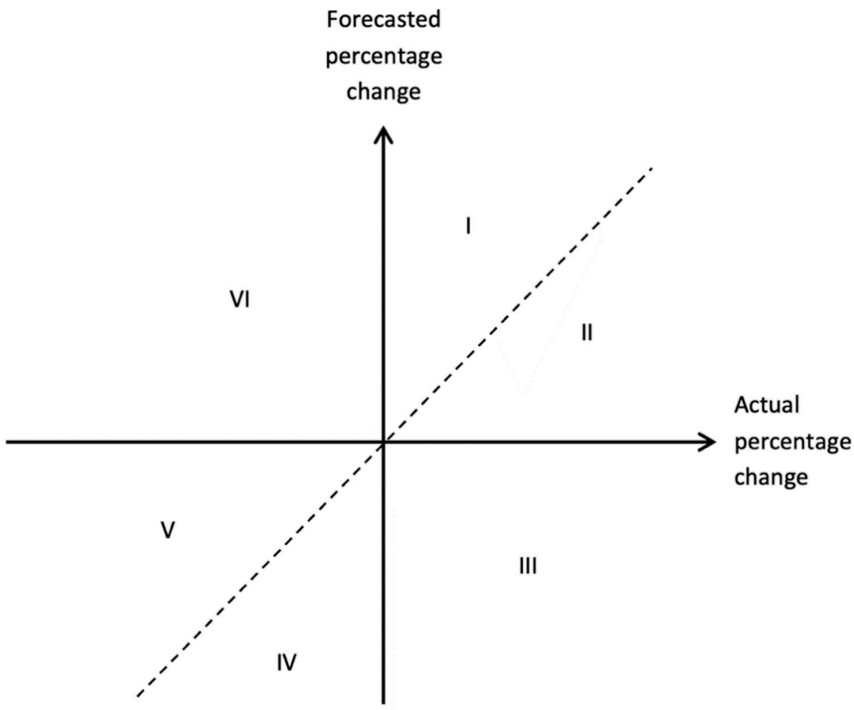


Figure 2. Prediction-realization diagram following Theil (1958). I. The percentage increase of the stock market index is overestimated. II. The percentage increase of the stock market index is underestimated. III. The stock market index rises, although a fall is forecasted. IV. The percentage decrease of the stock market index is overestimated. V. The percentage decrease of the stock market index is underestimated. VI. The stock market index falls, although a rise is forecasted.

Finally, we compare the forecasts with the naïve forecast. A forecaster who has obtained a notable insight into the future trend of the subject matter should at least be able to make more accurate forecasts than if one were to always assume that nothing at all will change (naïve forecast).

Simple measurements of forecast quality (such as the mean absolute squared error or the mean squared error) enable us to make a comparison with a naïve forecast. However, these simple approaches do not permit an assessment of statistical significance. This deficit is remedied by using the Diebold–Mariano test (Diebold and Mariano 1995). To do so, we calculate the mean squared error for the time series of the expert prognoses and for the time series of the naïve forecasts. The test statistics of the Diebold–Mariano test are defined as follows:

$$DM = \frac{\frac{1}{T}\sum(V(P_{t1}) - V(P_{t2}))}{\sqrt{\hat{\gamma}\, d/T}} \tag{3}$$

T = number of observations;
V = loss function;
P_1 = naïve forecast;
P_2 = expert forecast;
$\sqrt{\hat{\gamma}\, d/T}$ = joint spread of the two loss functions.

The null hypothesis tested in this way is that the naïve forecast (P_1) and the expert forecast (P_2) have the same accuracy. Neither one of the two alternatives thus provides clearly better results. The numerator is the mean deviation between the loss function V of the two forecasting approaches to be compared. Normally a squared loss function is assumed. In other words, the squared errors of the two forecast approaches are compared (P_1 and P_2). The denominator is the joint spread of the two loss functions. This is estimated on the basis of the long-term autocovariances of the loss function. In the case of large samples, this test value is asymptotically normally distributed.

As the methods and variables used by the forecasters in our data basis are not disclosed, we focus on the overall quality of the forecasts in terms of accuracy and unbiasedness. An assessment of the informative value of different forecasting approaches is not in the scope in this study.

5. Results

To provide a more detailed insight into our results, we first show the individual forecast quality of two selected German private banks. The graphic representation of the DAX forecasts of the German private bank Berenberg in a prediction-realization diagram indicates that conservative forecasting is at work here (Figure 3).

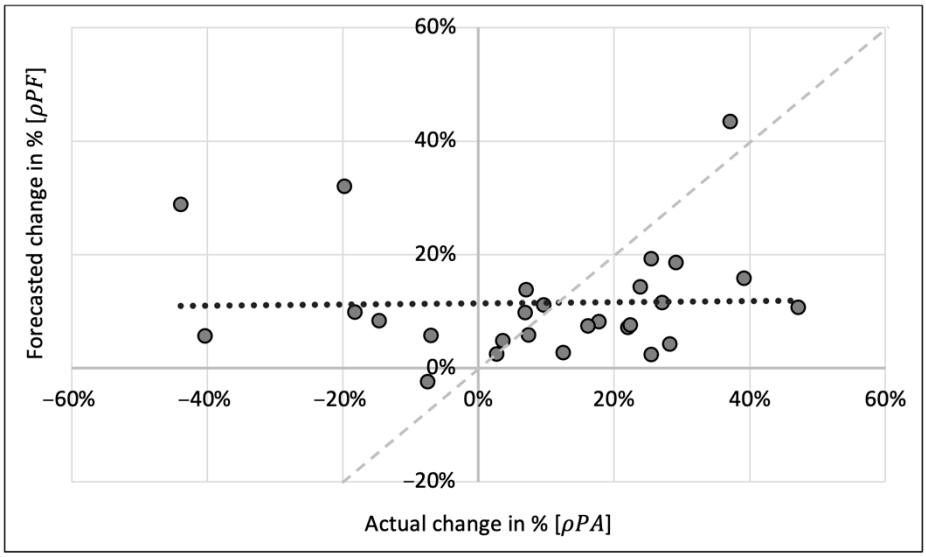


Figure 3. Prediction-realization diagram of the DAX forecasts of Berenberg. Dotted line = regression line; dashed line = perfect forecasts according to the prediction-realization diagram.

Berenberg issued a total of 27 DAX forecasts in the observation period (1992–2020). It is recognizable straight away that only one fall in the DAX is forecasted (3rd quadrant), but that the DAX actually does fall in seven out of the 27 years (3rd and 4th quadrant). This means that unusual events (falls in the DAX) are under-represented in the forecasts.

In addition, it can be seen that the dispersion of the actual events (scattering along the ρPA axis) is greater than the dispersion of the forecasts (scattering along the ρPF axis). The standard deviation of the actual events is 22.76%. The standard deviation of the forecasts, however, is only 9.98% (Table 3). The slope in the dotted regression line in the prediction-realization diagram of 0.011 is thus nowhere near the threshold value 1 (dashed diagonal line) (Table 3). The variability of the actual events is dramatically underestimated.

As another example, we consider the prediction-realization diagram of DAX forecasts made by the Franco-German private bank Oddo BHF (Figure 4).

This reveals a picture which is very similar to that of the prediction-realization diagram for Berenberg. In the period 1992–2020, at the end of each year Oddo BHF forecasts the DAX for the coming year. This occurs a total of 28 times. A fall in the DAX is forecasted on three occasions. In reality, however, the DAX falls in eight of the 28 years. This means that unusual events (falls in the DAX) are under-represented in the forecasts.

Table 3. The main results of the DAX forecasts from 1992 to 2020 from the *Handelsblatt*.

Institution	Forecasts Issued	Forecast		Actual		Normal Events over-Represented in the Forecasts	Standard Deviation		SD of the Forecasts < SD of the Actual Events	Regression Line		Slope of the Regression Lines < 1
		DAX Falls	DAX Rises	DAX Falls	DAX Rises		Forecast	Actual		Intercept	Slope	
Bank Julius Bär	23	2	21	8	15	Yes	0.062	0.248	Yes	0.088	−0.023	Yes
Bank of America	11	0	11	2	9	Yes	0.066	0.207	Yes	0.117	−0.001	Yes
Bankhaus Lampe	25	1	24	6	19	Yes	0.081	0.234	Yes	0.089	0.097	Yes
Bayerische Landesbank	26	1	25	6	20	Yes	0.067	0.230	Yes	0.080	−0.006	Yes
Berenberg	27	1	26	7	20	Yes	0.100	0.228	Yes	0.114	0.011	Yes
Bethmann Bank	12	2	10	5	7	Yes	0.095	0.284	Yes	0.101	−0.109	Yes
BNP Paribas	18	3	15	4	14	Yes	0.061	0.223	Yes	0.056	0.140	Yes
Commerzbank	28	2	26	7	21	Yes	0.089	0.234	Yes	0.120	−0.064	Yes
Credit Suisse	13	2	11	5	8	Yes	0.072	0.290	Yes	0.106	0.059	Yes
Dekabank	19	1	18	4	15	Yes	0.101	0.227	Yes	0.090	0.154	Yes
Deutsche Bank	25	2	23	7	18	Yes	0.070	0.237	Yes	0.091	−0.043	Yes
Dresdner Bank	15	0	15	5	10	Yes	0.084	0.276	Yes	0.080	0.099	Yes
DZ Bank	29	7	22	8	21	Yes	0.107	0.231	Yes	0.073	0.088	Yes
Haspa	13	0	13	3	10	Yes	0.047	0.202	Yes	0.080	0.045	Yes
Hauck & Aufhäuser	26	5	21	6	20	Yes	0.101	0.235	Yes	0.072	−0.040	Yes
Helaba	28	8	20	7	21	No	0.108	0.234	Yes	0.053	0.092	Yes
HSBC Trinkaus	22	3	19	7	15	Yes	0.085	0.256	Yes	0.080	−0.022	Yes
J.P. Morgan	22	4	18	6	16	Yes	0.100	0.244	Yes	0.084	0.038	Yes
LBB Landesbank Berlin	18	3	15	6	12	Yes	0.140	0.233	Yes	0.088	0.027	Yes
LBBW	21	1	20	6	15	Yes	0.107	0.226	Yes	0.090	0.093	Yes
Lehman Brothers	12	5	7	4	8	No	0.098	0.259	Yes	0.040	0.062	Yes
M.M. Warburg & Co.	29	3	26	8	21	Yes	0.091	0.231	Yes	0.076	−0.016	Yes
Morgan Stanley	14	6	8	4	10	No	0.123	0.285	Yes	0.030	0.136	Yes
National-Bank	15	3	12	3	12	No	0.086	0.202	Yes	0.082	0.028	Yes
NATIXIS	17	1	16	3	14	Yes	0.065	0.231	Yes	0.077	0.057	Yes
NordLB	12	2	10	2	10	No	0.038	0.153	Yes	0.041	−0.089	Yes
Oddo BHF	28	3	25	8	20	Yes	0.104	0.234	Yes	0.090	0.059	Yes
Pictet & Cie.	13	3	10	5	8	Yes	0.114	0.279	Yes	0.092	−0.074	Yes
Postbank	11	0	11	3	8	Yes	0.069	0.225	Yes	0.098	0.087	Yes
Sal. Oppenheim	21	2	19	5	16	Yes	0.093	0.248	Yes	0.067	0.111	Yes
Santander	24	1	23	7	17	Yes	0.093	0.239	Yes	0.116	0.101	Yes
Société Générale	20	4	16	5	15	Yes	0.096	0.228	Yes	0.065	0.043	Yes
SYZ & Co.	10	0	10	2	8	Yes	0.058	0.235	Yes	0.144	−0.042	Yes
UBS	14	3	11	4	10	Yes	0.120	0.242	Yes	0.112	0.007	Yes
Unicredit HypoVereinsbank	28	3	25	8	20	Yes	0.079	0.233	Yes	0.083	0.043	Yes
VP Bank	11	1	10	2	9	Yes	0.042	0.155	Yes	0.084	0.034	Yes
WestLB	21	3	18	7	14	Yes	0.106	0.260	Yes	0.081	0.124	Yes
WGZ Bank	16	1	15	5	11	Yes	0.172	0.211	Yes	0.110	0.301	Yes
Consensus	29	1	28	8	21	Yes	0.065	0.231	Yes	0.085	0.037	Yes
All forecasts	964	117	847	264	700	Yes	0.091	0.230	Yes	0.084	0.034	Yes

DAX = German Stock Market Index; SD = standard deviation.

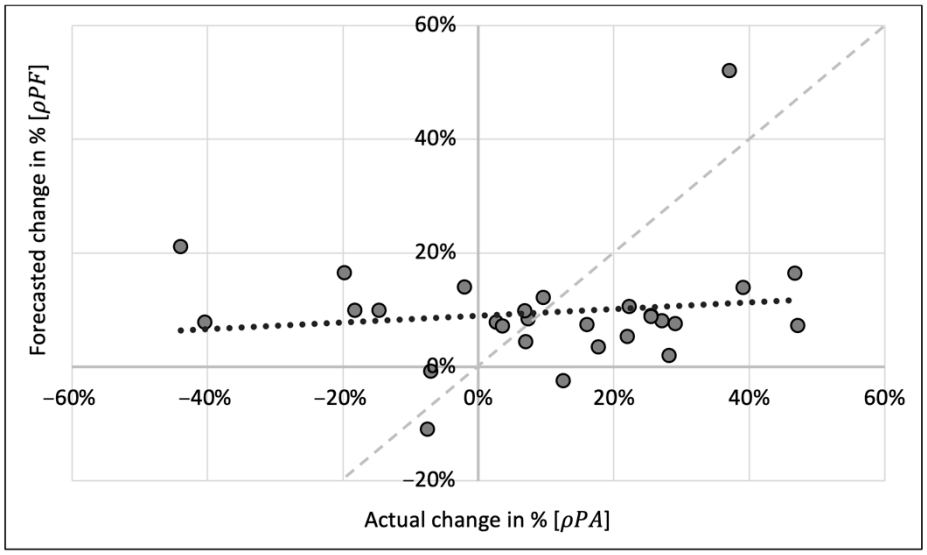


Figure 4. Prediction-realization diagram of the DAX forecasts of Oddo BHF. Dotted line = regression line; dashed line = perfect forecasts according to the prediction-realization diagram.

In addition, it can be seen that the dispersion of the actual events (scattering along the ρPA axis) is greater than the dispersion of the forecasts (scattering along the ρPF axis). The standard deviation of the actual events is 23.39%. The standard deviation of the forecasts, however, is only 10.41% (Table 3). The slope of 0.059 in the dotted regression line in the prediction-realization diagram is thus nowhere near the threshold value 1 (dashed diagonal line) (Table 3). The variability of the actual events is dramatically underestimated.

Table 3 depicts the main results of the DAX forecasts from the *Handelsblatt* newspaper. All of the forecasters who have taken part in the forecasting surveys of the *Handelsblatt* for at least ten years are analyzed individually. All of the forecasters who issue less than 10 forecasts in the period from 1992 to 2020 are not analyzed individually but are taken into account as part of the overall analysis of all forecasts and within the framework of the consensus forecasts (final lines in Table 3).

The seventh column of Table 3 indicates whether fewer falls in the DAX are forecasted than actually occur. As the DAX is a performance index and exhibits a rising trend over the long term, all falls in the index are interpreted as 'unusual events'. According to Ogburn (1934), conservative forecasting leads to 'normal events' (here: an increase in the DAX) being over-represented in the forecasts, while 'unusual events' (here: a decrease in the DAX) are under-represented in the forecasts. This is the case in 33 of the 38 forecasters who are analyzed individually here: a proportion of 86.8%. Unusual events are also under-represented in the consensus forecasts and when the total number of the forecasts is considered as a whole. The detailed data is given in Table 3, where one can see how often a falling DAX was forecast, and how often the DAX really falls. One can also note how often an upward trend was forecast for the DAX, and how often the DAX really rises (Table 3).

The picture is clearer in the case of the standard deviations. According to Ogburn (1934), conservative forecasting leads to standard deviations of the forecasts which are lower than the standard deviations of the actual events. The tenth column of Table 3 considers whether this applies to the DAX forecasts and reveals that this is the case in all 38 of the 38 forecasters analyzed. Also, with regard to the consensus forecasts and when all 964 forecasts are considered, the standard deviation of the forecasts lags behind the standard deviation of the actual events (Table 3).

Ogburn (1934) states that conservative forecasting leads to an underestimation of the variability of reality. In the prediction-realization diagram, this should lead to a slope in the regression lines which is lower than one. The last column of Table 3 illustrates this aspect. It can be seen that in 38 out of 38 cases, the slope in the regression lines is lower than one.

The fact that the slopes are usually clearly below the threshold value of one is also revealed in the detailed data on the intercepts and the slopes in the regression lines (Table 3).

The German quality newspaper the *Frankfurter Allgemeine Zeitung* (FAZ) only started a regular survey of forecasts in 2002. As a result, the share price falls in the years 2000 and 2001 no longer have an effect. It is interesting to see whether this leads to significantly different results in the forecasts. In addition, the *Frankfurter Allgemeine Zeitung* not only surveys annual forecasts, but also six-month forecasts. It is quite possible that the characteristics of the forecasts with differing forecast horizons vary. Once again, all of the forecasters who have taken part in the forecasting surveys of the FAZ at least ten times are analyzed individually (Table 4).

The results are in fact somewhat less clear than those for the DAX forecasts from the *Handelsblatt*. In 24 out of 33 cases (72.7%), normal events (increase in the DAX) are over-represented in the forecasts (seventh column in Table 4). Unusual events are also under-represented in the consensus forecasts and when all 282 six-month and all 402 twelve-month forecasts are considered as a whole.

The result of the standard deviations is quite clear: In 31 out of 33 cases (93.9%), the forecasts lag behind the actual events (tenth column in Table 4). This finding also applies to the consensus forecasts as well as when all 282 six-month and all 402 twelve-month forecasts are considered as a whole.

The fact that the forecasters persistently underestimate the variability of reality is revealed most clearly in the slope of the regression lines in the prediction-realization diagram (last column in Table 4). In 33 out of 33 cases, the slope is below one. This result also applies to the consensus forecasts as well as when all 282 six-month and all 402 twelve-month forecasts are considered as a whole.

The forecasts of the Dow Jones Industrial Index yield only slightly different results. Once again, all of the forecasters who have taken part in the forecasting survey at least ten times are analyzed individually (Table 5).

The Dow Jones Industrial Index is a price index, but it exhibits a long-term rising trend, nevertheless. To this extent, one can also presume here that a rise in the index can be considered a normal event, and that a fall in the index represents an unusual event. In ten out of 16 cases (62.5%), normal events (increase of the Dow Jones Industrial Index) are over-represented in the forecasts (seventh column in Table 5). Unusual events are also under-represented in the consensus forecasts and when all 203 six-month and all 259 twelve-month forecasts are considered as a whole.

The result for the standard deviations is more marked. In 14 out of 16 cases (87.5%), the fluctuations in the forecasts lag behind those of the actual events (tenth column in Table 5). This finding also applies to the consensus forecasts as well as when all 203 six-month and all 259 twelve-month forecasts are considered as a whole.

The fact that the forecasters persistently underestimate the variability of reality is revealed most clearly in the slope of the regression lines in the prediction-realization diagram (last column in Table 5). In 16 out of 16 cases, the slope is below one. This is also the same for the consensus forecasts as well as when all 203 six-month and all 259 twelve-month forecasts are viewed as a whole.

The picture drawn by the forecasts of the Euro Stoxx 50 is even more distinct (Table 6). Here again, all of the forecasters who have taken part in the forecasting survey at least ten times are analyzed individually. All of the other forecasts form part of the consensus forecasts and are also evaluated as part of the total number of forecasts.

Table 4. The main results of the DAX forecasts from 2002 to 2020 from the FAZ.

Institution	Forecasts Issued	Forecast		Actual		Normal Events over-Represented in the Forecasts	Standard Deviation		SD of the Forecasts < SD of the Actual Events	Regression Line		Slope of the Regression Lines < 1
		DAX Falls	DAX Rises	DAX Falls	DAX Rises		Forecast	Actual		Intercept	Slope	
Forecast horizon 6 months												
Bayern LB	10	5	5	3	7	No	0.047	0.094	Yes	0.028	−0.286	Yes
Deka Bank	16	3	13	5	11	Yes	0.061	0.096	Yes	0.040	−0.002	Yes
DZ Bank	16	6	10	5	11	No	0.065	0.096	Yes	0.009	0.032	Yes
Helaba	14	6	8	5	9	No	0.075	0.102	Yes	0.025	−0.375	Yes
HSH Nordbank	10	7	3	4	6	No	0.095	0.098	Yes	−0.030	−0.039	Yes
HVB-Unicredit Bank	16	4	12	6	10	Yes	0.063	0.104	Yes	0.035	−0.035	Yes
LBBW	17	3	14	6	11	Yes	0.048	0.102	Yes	0.019	0.090	Yes
M.M. Warburg	17	3	14	6	11	Yes	0.122	0.102	No	0.030	−0.039	Yes
Oddo BHF	10	1	9	4	6	Yes	0.041	0.121	Yes	0.049	−0.058	Yes
Postbank	13	6	7	4	9	No	0.071	0.104	Yes	0.008	−0.087	Yes
Santander Asset Mgmt.	13	1	12	3	10	Yes	0.029	0.099	Yes	0.033	0.073	Yes
Société Générale	10	6	4	3	7	No	0.087	0.072	No	−0.023	−0.431	Yes
Consensus	17	2	15	6	11	Yes	0.028	0.102	Yes	0.024	−0.077	Yes
All forecasts	282	83	199	103	179	Yes	0.072	0.095	Yes	0.024	−0.076	Yes
Forecast horizon 12 months												
Allianz SE	11	0	11	2	9	Yes	0.044	0.155	Yes	0.072	0.018	Yes
Bayern LB	11	0	11	2	9	Yes	0.036	0.159	Yes	0.069	0.011	Yes
BNP Paribas	12	1	11	3	9	Yes	0.055	0.210	Yes	0.066	0.110	Yes
Commerzbank	18	0	18	4	14	Yes	0.081	0.233	Yes	0.119	0.032	Yes
Deka Bank	18	1	17	3	15	Yes	0.104	0.195	Yes	0.082	0.200	Yes
Deutsche Bank	10	0	10	2	8	Yes	0.047	0.212	Yes	0.104	−0.017	Yes
DWS	13	0	13	3	10	Yes	0.027	0.202	Yes	0.076	0.038	Yes
DZ Bank	18	2	16	4	14	Yes	0.066	0.222	Yes	0.072	0.063	Yes
Helaba	15	6	9	3	12	No	0.121	0.196	Yes	0.025	0.249	Yes
HSBC Trinkaus & Burkhardt	13	2	11	3	10	Yes	0.066	0.262	Yes	0.065	−0.102	Yes
HSH Nordbank	11	2	9	3	8	Yes	0.080	0.213	Yes	0.055	0.192	Yes
HVB-Unicredit Bank	18	1	17	4	14	Yes	0.078	0.228	Yes	0.077	0.077	Yes
J.P. Morgan	12	1	11	3	9	Yes	0.064	0.233	Yes	0.095	0.140	Yes
LBBW	19	0	19	4	15	Yes	0.097	0.227	Yes	0.091	0.093	Yes
M.M. Warburg	19	1	18	4	15	Yes	0.097	0.227	Yes	0.078	−0.018	Yes
Oddo BHF	17	1	16	4	13	Yes	0.045	0.225	Yes	0.093	−0.092	Yes
Postbank	14	0	14	3	11	Yes	0.070	0.208	Yes	0.096	0.048	Yes
Santander Asset Mgmt.	16	0	16	3	13	Yes	0.052	0.195	Yes	0.107	0.048	Yes
Société Générale	11	4	7	2	9	No	0.088	0.155	Yes	0.067	−0.347	Yes
UBS	10	1	9	1	9	No	0.118	0.151	Yes	0.136	0.027	Yes
WestLB	11	2	9	3	8	Yes	0.128	0.282	Yes	0.075	0.204	Yes
Consensus	19	0	19	4	15	Yes	0.061	0.227	Yes	0.087	0.064	Yes
All forecasts	402	31	371	88	314	Yes	0.083	0.215	Yes	0.085	0.054	Yes

DAX = German Stock Market Index; FAZ = Frankfurter Allgemeine Zeitung; SD = standard deviation.

Table 5. Main results of the forecasts of the DJI from 2004 to 2020 from the FAZ.

Institution	Forecasts Issued	Forecast		Actual		Normal Events over-Represented in the Forecasts	Standard Deviation		SD of the Forecasts < SD of the Actual Events	Regression Line		Slope of the Regression Lines < 1
		DJI Falls	DJI Rises	DJI Falls	DJI Rises		Forecast	Actual		Intercept	Slope	
Forecast horizon 6 months												
Deka Bank	15	5	10	8	7	Yes	0.070	0.066	No	0.018	0.171	Yes
Helaba	14	6	8	6	8	No	0.081	0.077	No	0.019	−0.406	Yes
LBBW	16	7	9	8	8	Yes	0.052	0.073	Yes	0.010	0.116	Yes
M.M. Warburg	15	3	12	7	8	Yes	0.061	0.075	Yes	0.034	0.233	Yes
Postbank	12	6	6	5	7	No	0.053	0.079	Yes	0.003	0.035	Yes
Santander Asset Mgmt.	13	1	12	6	7	Yes	0.019	0.081	Yes	0.026	−0.095	Yes
Consensus	16	4	12	8	8	Yes	0.019	0.073	Yes	0.014	0.036	Yes
All forecasts	203	67	136	106	97	Yes	0.061	0.070	Yes	0.014	0.040	Yes
Forecast horizon 12 months												
BNP Paribas	10	0	10	3	7	Yes	0.040	0.183	Yes	0.072	−0.059	Yes
Commerzbank	10	0	10	3	7	Yes	0.052	0.169	Yes	0.081	0.120	Yes
Deka Bank	16	6	10	4	12	No	0.099	0.137	Yes	0.051	0.002	Yes
Helaba	15	7	8	3	12	No	0.107	0.149	Yes	0.008	0.193	Yes
HSH Nordbank	11	5	6	3	8	No	0.067	0.163	Yes	0.022	−0.032	Yes
LBBW	17	4	13	4	13	No	0.058	0.142	Yes	0.053	−0.042	Yes
M.M. Warburg	17	1	16	4	13	Yes	0.071	0.142	Yes	0.063	−0.107	Yes
Oddo BHF	15	0	15	3	12	Yes	0.022	0.147	Yes	0.058	0.054	Yes
Postbank	13	0	13	3	10	Yes	0.063	0.160	Yes	0.084	0.012	Yes
Santander Asset Mgmt.	16	0	16	4	12	Yes	0.051	0.146	Yes	0.070	0.093	Yes
Consensus	17	0	17	4	13	Yes	0.033	0.142	Yes	0.055	0.006	Yes
All forecasts	259	33	226	65	194	Yes	0.066	0.140	Yes	0.057	0.029	Yes

DJI = Dow Jones Industrial Index; FAZ = Frankfurter Allgemeine Zeitung; SD = standard deviation.

Table 6. The main results for the Euro Stoxx 50 forecasts from 2002 to 2020 from the FAZ.

Institution	Forecasts Issued	Forecast		Actual		Normal Events over-Represented in the Forecasts	Standard Deviation		SD of the Forecasts < SD of the Actual Events	Regression Line		Slope of the Regression Lines < 1
		SX5E Falls	SX5E Rises	SX5E Falls	SX5E Rises		Forecast	Actual		Intercept	Slope	
Forecast horizon 6 months												
Bayern LB	10	4	6	5	5	Yes	0.043	0.078	Yes	0.011	−0.244	Yes
Deka Bank	16	3	13	8	8	Yes	0.063	0.093	Yes	0.049	0.022	Yes
DZ Bank	16	3	13	8	8	Yes	0.064	0.093	Yes	0.030	0.186	Yes
Helaba	14	6	8	8	6	Yes	0.079	0.095	Yes	0.019	−0.406	Yes
HSH Nordbank	10	6	4	6	4	No	0.085	0.099	Yes	−0.030	−0.214	Yes
HVB-Unicredit Bank	16	3	13	8	8	Yes	0.070	0.101	Yes	0.023	−0.085	Yes
LBBW	17	6	11	9	8	Yes	0.053	0.098	Yes	0.028	0.088	Yes
M.M. Warburg	16	2	14	8	8	Yes	0.073	0.101	Yes	0.055	−0.014	Yes
Oddo BHF	10	2	8	5	5	Yes	0.042	0.116	Yes	0.033	−0.009	Yes
Postbank	13	6	7	7	6	Yes	0.060	0.097	Yes	0.004	−0.100	Yes
Santander Asset Mgmt.	13	2	11	6	7	Yes	0.033	0.099	Yes	0.030	0.110	Yes
Consensus	17	5	12	9	8	Yes	0.030	0.098	Yes	0.023	−0.018	Yes
All forecasts	270	82	188	144	126	Yes	0.073	0.094	Yes	0.023	−0.007	Yes
Forecast horizon 12 months												
Allianz SE	11	0	11	4	7	Yes	0.042	0.130	Yes	0.071	−0.035	Yes
Bayern LB	11	0	11	3	8	Yes	0.039	0.127	Yes	0.058	−0.044	Yes
BNP Paribas	11	1	10	3	8	Yes	0.044	0.194	Yes	0.076	−0.069	Yes
Commerzbank	18	1	17	5	13	Yes	0.064	0.195	Yes	0.080	0.017	Yes
Deka Bank	18	1	17	5	13	Yes	0.093	0.170	Yes	0.094	0.107	Yes
DWS	12	0	12	5	7	Yes	0.043	0.175	Yes	0.078	−0.019	Yes
DZ Bank	18	1	17	6	12	Yes	0.075	0.193	Yes	0.090	0.096	Yes
Helaba	15	5	10	5	10	No	0.117	0.177	Yes	0.048	0.292	Yes
HSBC Trinkaus&Burkhardt	14	3	11	4	10	Yes	0.082	0.209	Yes	0.065	−0.141	Yes
HSH Nordbank	11	1	10	4	7	Yes	0.071	0.195	Yes	0.076	0.119	Yes
HVB-Unicredit Bank	18	0	18	6	12	Yes	0.064	0.193	Yes	0.070	0.050	Yes
LBBW	19	1	18	6	13	Yes	0.078	0.190	Yes	0.088	0.003	Yes
M.M. Warburg	19	1	18	6	13	Yes	0.083	0.190	Yes	0.074	−0.073	Yes
Oddo BHF	17	1	16	6	11	Yes	0.047	0.192	Yes	0.072	−0.074	Yes
Postbank	14	0	14	4	10	Yes	0.054	0.190	Yes	0.086	0.032	Yes
Santander Asset Mgmt.	16	0	16	5	11	Yes	0.053	0.178	Yes	0.095	0.078	Yes
WestLB	11	1	10	4	7	Yes	0.088	0.231	Yes	0.073	0.127	Yes
Consensus	19	0	19	6	13	Yes	0.044	0.190	Yes	0.083	0.020	Yes
All forecasts	381	29	352	123	258	Yes	0.073	0.179	Yes	0.080	0.017	Yes

SX5E = Euro Stoxx 50; FAZ = Frankfurter Allgemeine Zeitung; SD = standard deviation.

Conservatism among forecasters can lead to them forecasting unusual events too rarely. The Euro Stoxx 50 is a price index, but in spite of this it exhibits a long-term upward trend. To this extent, one can also presume here that a rise in the index can be considered a normal event, and that a fall in the index represents an unusual event. In the predictions of 24 of the 26 forecasters analyzed individually (92.3%), unusual events are under-represented (seventh column in Table 6). The consensus forecasts and the overall total of all 270 six-month forecasts and all 381 twelve-month forecasts also show that unusual events are forecast more seldom than they occur in reality.

The standard deviations provide a very clear picture. The standard deviations of the forecasts lag behind the standard deviations of the actual results in 26 out of 26 cases (tenth column in Table 6). This also applies to the consensus forecasts and the overall total of 270 forecasts with a forecast horizon of six months and all 381 forecasts with a forecast horizon of twelve months.

Finally, it can be seen that the slope in the regression lines in the prediction-realization diagrams is significantly below one in 26 out of 26 cases. The forecasters are thus obviously underestimating the variability of reality (last column in Table 6). These findings are also confirmed when the consensus forecasts and the overall total number of forecasts are considered.

Without exception, it can be observed that the forecasters underestimate the variability of reality. This fact can be clearly seen when looking at the kernel density plots of the forecast relative changes in share prices and the actual relative changes in share prices (Figure 5). The spread of the forecasts is much smaller than the spread of the actual events.

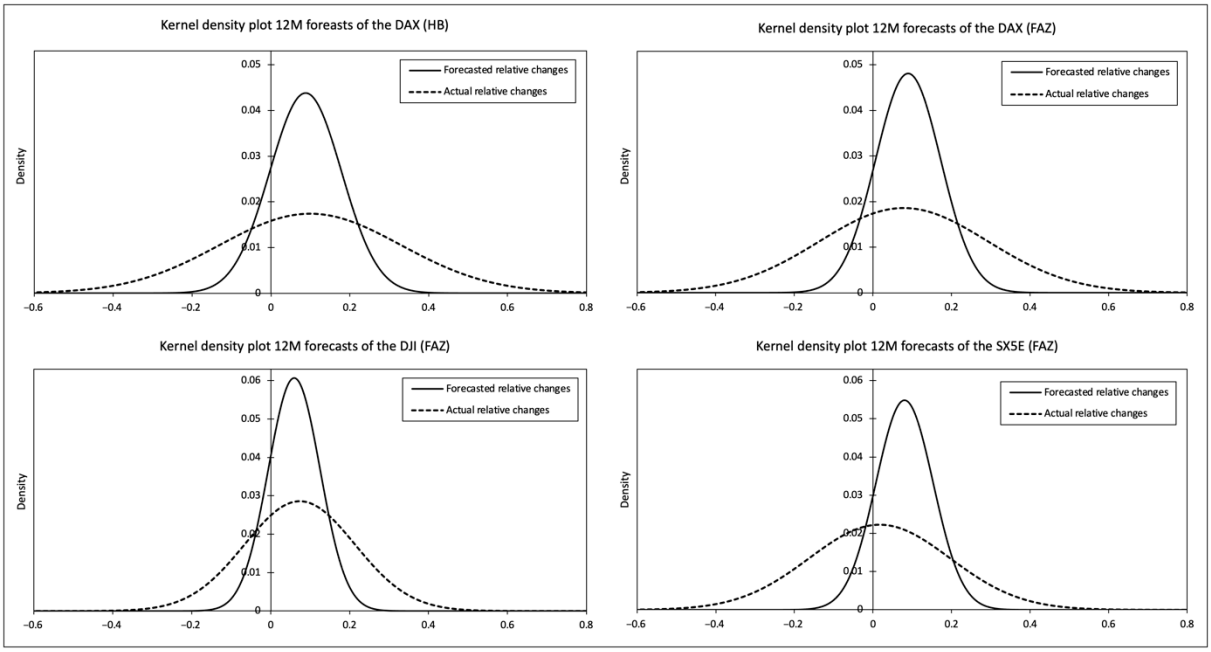


Figure 5. Kernel density plots of the forecast and actual changes of stock-market indices. HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; DAX = German Stock Market Index; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; 12M = Forecast horizon of 12 months.

This can be also seen in the fact that the slope in the regression lines in the prediction-realization diagram always remains below the threshold value of one. This leads us to the assessment that this aspect in particular deserves special attention. The unbiasedness test takes the slope of the regression line in the prediction-realization diagram into account as an essential element. Forecasts are viewed as unbiased when the slope in the regression line does not diverge significantly from one, the intercept of the regression line does not deviate significantly from zero, and the residuals are randomly distributed. The decisive

advantage of this approach lies in the opportunity to go beyond purely descriptive statistics and to examine the statistical significance of the results.

In all seven cases, it can be seen that given an error probability of ≤1% either the slope of the regression line in the prediction-realization diagram is ≠1 and/or the intercept is ≠ 0. In addition, the residuals are obviously not randomly distributed in six of the seven cases. The forecasts are clearly not unbiased (Table 7).

Table 7. Unbiasedness test.

Stock Market Index	Source	Forecast Horizon	Number of Observations	Slope	Intercept	F Test *p*-Value	Wooldridge Test *p*-Value
DAX	HB	12M	964	0.034	0.084	0.000	0.000
DAX	FAZ	6M	282	−0.075	0.024	0.000	0.000
DAX	FAZ	12M	402	0.054	0.085	0.000	0.006
DJI	FAZ	6M	203	0.040	0.014	0.010	0.098
DJI	FAZ	12M	259	0.029	0.057	0.000	0.623
SX5E	FAZ	6M	270	−0.007	0.023	0.000	0.091
SX5E	FAZ	12M	381	0.017	0.080	0.000	0.042

DAX = German Stock Market Index; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; 12M = 12 months; 6M = 6 months.

Finally, with the aid of the Diebold–Mariano test we examine whether the quality of the forecasts is significantly superior—from a statistical perspective—to that of naïve forecasts (Table 8). The result is that the forecasts of the Euro Stoxx 50 are significantly poorer than the corresponding naïve forecasts, and the quality of the forecasts for the DAX and the Dow Jones Industrial Index does not go significantly beyond that of naïve forecasts.

Table 8. Comparison of the forecasts with the naïve forecast.

Stock Market Index	Source	Forecast Horizon	Diebold–Mariano Test	
			Result	*p*-Value
DAX	HB	12M	o	0.8143
DAX	FAZ	6M	o	0.1221
DAX	FAZ	12M	o	0.7429
DJI	FAZ	6M	o	0.7053
DJI	FAZ	12M	o	0.3491
SX5E	FAZ	6M	−	0.0000
SX5E	FAZ	12M	−	0.0540

o = no significant result, − = significantly poorer than the naïve forecasts, + = significantly better than the naïve forecast, DAX = German Stock Market Index; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; 12M = 12 months; 6M = 6 months.

In Table 9 the results of the hypothesis testing are summarized. In Hypotheses 1–3, the result which was determined for "all forecasts" in a forecasting area is used. In the case of the DAX forecasts from the *Handelsblatt* survey, for example, that is the 964 forecasts which are noted in the final line of Table 3. For Hypothesis 4, the results of the unbiasedness test (Table 7) are taken into account, and for Hypothesis 5 the results of the Diebold–Mariano test (Table 8).

In the case of Hypothesis 1 there is a uniform pattern for all areas of forecasting and all forecast horizons. Normal events (index rises) are over-represented in the forecasts. Unusual events (index falls) are under-represented in the forecasts. Null Hypothesis 1 has to be rejected in all seven cases.

Table 9. The results of hypothesis testing.

Stock Market Index	Source	Forecast Horizon	Hypothesis 1	Hypothesis 2	Hypothesis 3	Hypothesis 4	Hypothesis 5
DAX	HB	12M	+	+	+	+	+
DAX	FAZ	6M	+	+	+	+	+
DAX	FAZ	12M	+	+	+	+	+
DJI	FAZ	6M	+	+	+	+	+
DJI	FAZ	12M	+	+	+	+	+
SX5E	FAZ	6M	+	+	+	+	+
SX5E	FAZ	12M	+	+	+	+	+

+ = Null Hypothesis rejected; DAX = German Stock Market Index; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; 12M = 12 months; 6M = 6 months.

In the case of Hypothesis 2 there are no differences between the subjects of the forecasts and the forecast horizons. In all seven cases, Null Hypothesis 2 has to be rejected. The dispersion of the forecasts (measured against the standard deviation) thus lags behind the dispersion of the actual events.

A uniform picture is also shown with regard to Hypothesis 3. In all seven forecasting areas the slope of the regression line in the prediction-realization diagrams is clearly below one. Null Hypothesis 3 has to be rejected in all seven cases. This means that the rates of change of the stock-market indices are significantly underestimated.

In the case of Hypothesis 4 there are also no relevant differences regarding the subjects of the forecasts or the forecast horizons. In all seven areas, the forecasts prove to be biased. These results are highly significant. In all seven cases, Null Hypothesis 4 has to be rejected.

In Hypothesis 5 there is also a concurring result for all seven forecast groups. Null Hypothesis 5 has to be discarded. The precision of the forecasts does not go significantly beyond that of naïve forecasts.

The findings of Ogburn (1934) are thus fully confirmed in the stock market forecasts which we analyzed. It can certainly be stated that these stock-market analysts systematically underestimate the variability of reality and that the success rate of their forecasts does not extend beyond that of naïve forecasts. Their behavior can be described as "sticky" because their forecasts adhere too strongly to long-term trends in the indices to provide meaningful information about current events.

This study expands on existing research as it is the first of its kind to analyze ex-ante forecasts for the SX5E. The picture obtained is similar to that of the stock indices examined previously. The forecasts are mostly biased and not significantly better than naïve forecasts. About 15 years ago, ex-ante forecasts for the DAX and the DJI were last examined (Table 1). In the meantime, technological progress has led to the emergence of numerous promising new forecasting methods, as discussed in our literature review. However, our results indicate that this has not, at least so far, contributed to a significant increase in the quality of the forecast.

Our findings allow different conclusions to be drawn with regard to the efficient market hypothesis (Fama 1970). On the one hand, the Diebold–Mariano test shows that the forecast quality is poor. This is compatible with the efficient market hypothesis, since no excess returns can be achieved on the basis of the forecasts. On the other hand, the efficient market hypothesis assumes that economic subjects are fully informed. The permanent underestimation of the variability of reality that the prediction-realization diagram reveals should therefore not occur. The acting subjects do not seem to take notice of the discrepancy between their own actions and reality, since no correction of the behavior is made in the subsequent forecasts.

The forecasters systematically underestimate the variability of reality. Against the background of Mandelbrot's fractal theory, it seems reasonable to conclude that forecasters—as long as they think in terms of "trending" and "mean reversion"—systematically underestimate the Hurst exponent (Mandelbrot 2004) of stock market developments.

Overall, the forecast quality for all three indices is not sufficient to enable an active investment strategy on the basis of the forecasts that is likely to be successful. Moreover, since unusual events (e.g., a sudden drop in an otherwise rising trendline) are seldom successfully forecasted, an active investment strategy based on the forecasts harbors risks that can cause severe financial damage to investors. Thus, we advise private and professional investors to consider a passive investment strategy instead when deciding how to invest their assets.

The path which has to be followed to obtain better stock market forecasts thus becomes clear: analysts have to be more courageous. They need to react to new trends with more flexibility. They have to leave their comfort zone more frequently and stand by assessments which are not necessarily approved of by the majority of their peers. That alone will presumably not suffice to generate reliable stock market forecasts: they will also need to work hard on the quality of their approaches to forecasting. To this end, a variety of interesting approaches are already discussed in the literature, e.g., economic forecasts based on newspaper texts or news from online media and attention to news events (Milas et al. 2021; Kalamara et al. 2020; Ben-Rephael et al. 2017). If analysts want to significantly improve the reliability of their forecasts, there is no alternative but to change their overly cautious, highly conservative, and thus inflexible attitudes.

Finally, our study also has some limitations. First of all, it should be mentioned that we are looking at forecasts for entire stock indices. Even if the forecasters do not manage to successfully predict the development of a stock index, this does not mean that the entire stock market is per se unsuitable for an active investment strategy. It is still conceivable that stocks of individual companies in the index can be predicted successfully. In this case, an active investment strategy based on the forecasts for individual stocks could be very promising. Second, forecasting future events with a six- to twelve-month horizon is a major hurdle. As the forecast quality tends to increase as the horizon decreases (Dua 1988), it is conceivable that, for example, monthly forecasts for the same indices would lead to significantly better results. Last but not least, we analyze the entire time series from beginning to end for each forecaster. Even though this leads to a large sample size, which enables a clearer picture of the forecast quality overall, differences in the forecast quality over time may remain undetected. This could be the case in particular for the forecasts published in *Handelsblatt*, which extend over a period of 29 years.

Our results provide initial indications that patterns discovered almost 90 years ago that massively deteriorate forecast quality can still be found in stock market forecasts today. We therefore encourage future research efforts to examine whether our results prevail in additional datasets. Furthermore, we believe that deeper analysis of the rationale for conservative forecasting and an assessment of its financial impact on investors are promising areas of research that would deepen our understanding of ex-ante stock market forecasts.

6. Summary

We examine forecasts for the German Stock Market Index (DAX), the Dow Jones Industrial Index (DJI), and the Euro Stoxx 50 (SX5E) which were published in the period 1992 to 2020 in the German business newspaper *Handelsblatt* (HB) and the quality broadsheet the *Frankfurter Allgemeine Zeitung* (FAZ). These forecasts have a horizon of six and twelve months. The forecasts are from German and international banks such as Deutsche Bank, Goldman Sachs, J.P. Morgan, or BNP Paribas.

We take up the thoughts of Ogburn (1934), who, on the basis of a small empirical survey, became convinced that forecasters consistently underestimate the variability of the future, and that their forecasting is of a conservative nature. However, we also go beyond this and use some contemporary measures (prediction-realization diagram, test of unbiasedness, Diebold–Mariano test) to test ex-ante forecasts for their success at the time of validity.

Conservative forecasting behavior leads to unusual events being under-represented in forecasts, to the dispersion of the forecasts (as measured by their standard deviation) lagging behind the dispersion of the actual events, and to the extent of the forecasted changes being smaller than the actual changes. The latter aspect is reflected in a flat course of the regression line in the prediction-realization diagram (slope < 1) and thus also leads to failure in the unbiasedness test.

We analyze a total of 2,761 forecasts which are divided up into seven groups according to the subject of the forecast (DAX, DJI, SX5E), the forecast horizon (6 and 12 months), and the source (FAZ, HB). The findings are that in all seven groups (a) unusual events are under-represented in the forecasts, (b) the dispersion of the forecasts lags behind that of actual events, (c) the slope in the regression lines in the prediction-realization diagram is <1, (d) the forecasts are biased to a highly significant degree, and (e) that the quality of the forecasts is not significantly better than that of naïve forecasts.

It is more than surprising how closely these stock market forecasts for the years 1992 to 2020 correspond to the characteristics which Ogburn described back in the 1930s. The stock market analysts prove to be too conservative, inflexible, and cautious. If they want to improve the reliability of their forecasts, they should change their conservative and inflexible forecasting behavior and consider promising new approaches and technologies in their forecasting process. For private and professional investors, building active investment strategies based on the insufficient stock market forecasts examined can involve enormous financial risks and is therefore not recommended.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/jrfm14120593/s1.

Author Contributions: Data curation, I.F., J.R.J., M.L. and M.S.; Formal analysis, I.F., J.R.J., M.L. and M.S.; Software, I.F.; Visualization, I.F., J.R.J. and M.L.; Writing—original draft, J.R.J., M.L. and M.S.; Writing—review & editing, I.F., J.R.J., M.L. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Our data basis can be accessed at Supplementary Materials.

Acknowledgments: The authors thank the editor and the anonymous reviewers for their constructive comments and useful suggestions, which were very helpful to enhance the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Detailed Summary Statistics on Data Basis

Table A1. Detailed Summary Statistics on DAX, DJI, and SX5E forecasts of our data basis.

Source	Subject	Year	N	Min [pts.]	Max [pts.]	Median [pts.]	Mean [pts.]	Actual [pts.]	N	Min [pts.]	Max [pts.]	Median [pts.]	Mean [pts.]	Actual [pts.]
					Forecast horizon 6 months						Forecast horizon 12 months			
HB	DAX	1992	NA	NA	NA	NA	NA	NA	21	1600	1900	1780	1764	1545.05
		1993	NA	NA	NA	NA	NA	NA	25	1550	1900	1750	1726	2266.68
		1994	NA	NA	NA	NA	NA	NA	28	1840	2500	2400	2339	2106.58
		1995	NA	NA	NA	NA	NA	NA	33	1950	2500	2200	2225	2253.88
		1996	NA	NA	NA	NA	NA	NA	28	2250	2700	2450	2449	2888.69
		1997	NA	NA	NA	NA	NA	NA	34	2600	3800	3100	3095	4249.69
		1998	NA	NA	NA	NA	NA	NA	33	4000	4800	4413	4413	5002.39
		1999	NA	NA	NA	NA	NA	NA	34	4580	6000	5400	5390	6958.14
		2000	NA	NA	NA	NA	NA	NA	37	6200	7620	6790	6771	6433.61
		2001	NA	NA	NA	NA	NA	NA	33	6100	9000	7800	7722	5160.10
		2002	NA	NA	NA	NA	NA	NA	38	5100	6650	5750	5779	2892.63
		2003	NA	NA	NA	NA	NA	NA	33	3300	5000	3915	3921	3965.16
		2004	NA	NA	NA	NA	NA	NA	34	3500	5000	4300	4318	4256.08
		2005	NA	NA	NA	NA	NA	NA	33	4100	5000	4600	4558	5408.26
		2006	NA	NA	NA	NA	NA	NA	38	5000	6100	5800	5717	6596.92
		2007	NA	NA	NA	NA	NA	NA	37	6000	7500	7078	7027	8067.32
		2008	NA	NA	NA	NA	NA	NA	35	7700	9250	8500	8566	4810.20
		2009	NA	NA	NA	NA	NA	NA	31	3600	6500	5250	5230	5957.43
		2010	NA	NA	NA	NA	NA	NA	38	4500	7500	6345	6339	6914.19
		2011	NA	NA	NA	NA	NA	NA	39	6200	8300	7600	7605	5898.35
		2012	NA	NA	NA	NA	NA	NA	37	5500	7600	6573	6573	7612.39
		2013	NA	NA	NA	NA	NA	NA	35	6900	8890	8029	8024	9552.16
		2014	NA	NA	NA	NA	NA	NA	33	8900	11,000	10,200	10,123	9805.55
		2015	NA	NA	NA	NA	NA	NA	36	9500	11,800	10,753	10,706	10,743.01
		2016	NA	NA	NA	NA	NA	NA	36	9250	13,000	11,850	11,793	11,481.06
		2017	NA	NA	NA	NA	NA	NA	30	11,000	12,300	11,800	11,724	12,917.64
		2018	NA	NA	NA	NA	NA	NA	33	12,300	15,000	14,000	14,009	10,558.96
		2019	NA	NA	NA	NA	NA	NA	31	10,000	13,400	12,000	12,053	13,249.01
		2020	NA	NA	NA	NA	NA	NA	31	12,500	15,000	14,000	13,999	13,718.78
					Forecast horizon 6 months						Forecast horizon 12 months			
FAZ	DAX	2002	14	4900	6000	5650	5554	4382.56	19	5100	6650	5750	5808	2892.63
		2003	NA	NA	NA	NA	NA	3220.58	17	3000	4200	3800	3780	3965.16
		2004	14	3600	4500	4200	4184	4052.73	15	3833	4700	4300	4299	4256.08
		2005	15	3900	4600	4400	4330	4586.28	21	4100	4750	4570	4560	5408.26
		2006	17	5000	5950	5700	5616	5683.31	20	5100	6100	5725	5689	6596.92
		2007	14	6200	7100	6612	6623	8007.32	20	6000	7400	7000	6988	8067.32
		2008	14	7250	8700	8066	8081	6418.32	18	7700	9200	8500	8503	4810.20
		2009	17	3200	5700	4900	4725	4808.64	17	3600	6500	5400	5353	5957.43
		2010	19	4800	6800	6000	5875	5965.52	22	5300	7100	6375	6333	6914.19
		2011	19	6300	8000	7300	7289	7376.24	26	6200	8300	7600	7618	5898.35
		2012	14	4800	7000	6105	6009	6416.28	22	5500	7600	6594	6588	7612.39
		2013	14	7000	8200	7659	7618	7959.22	20	7250	8890	8035	8069	9552.16
		2014	16	8500	10,200	9660	9620	9833.07	23	8900	11,000	10,150	10,092	9805.55
		2015	18	8700	11,000	10,300	10,035	10,944.97	23	9500	11,500	10,900	10,773	10,743.01
		2016	17	10,200	12,250	11,400	11,388	9680.09	23	10,800	12,600	11,900	11,859	11,481.06
		2017	19	10,600	12,400	11,500	11,494	12,325.12	24	10,400	12,300	11,800	11,713	12,917.64
		2018	19	12,500	15,000	13,700	13,658	12,306.00	25	12,300	14,500	14,000	13,938	10,558.96
		2019	NA	NA	NA	NA	NA	12,398.80	24	10,000	13,400	12,000	11,986	13,249.01
		2020	22	12,000	14,500	13,625	13,460	12,310.93	23	12,500	14,500	14,000	13,833	13,718.78
					Forecast horizon 6 months						Forecast horizon 12 months			
FAZ	DJI	2004	10	9800	11,000	10,422	10,444	10,435.48	10	10,000	11,200	10,500	10,544	10,783.01
		2005	10	10,800	11,200	11,010	11,020	10,274.97	14	11,000	12,000	11,420	11,440	10,717.50
		2006	14	10,000	11,800	11,223	11,196	11,150.22	15	10,300	12,500	11,500	11,575	12,463.15
		2007	12	12,200	14,000	12,800	12,805	13,408.62	14	11,440	14,000	13,400	13,276	13,264.82
		2008	13	12,500	14,500	13,729	13,729	11,350.01	16	13,500	15,300	14,500	14,513	8776.39
		2009	14	6900	10,800	9000	9000	8447.00	16	7000	12,500	9940	9880	10,428.05
		2010	16	8900	12,100	10,600	10,433	9774.02	18	10,000	12,100	11,050	11,118	11,577.51
		2011	14	10,500	13,900	11,904	11,808	12,414.34	16	10,200	13,500	12,064	12,127	12,217.56
		2012	9	10,800	13,500	12,363	12,363	12,880.09	13	12,375	15,000	13,200	13,240	13,104.14
		2013	8	12,100	14,000	13,487	13,381	14,909.60	11	13,000	15,300	14,150	14,150	16,576.66
		2014	12	14,500	16,800	16,500	16,364	16,826.60	14	15,700	17,700	17,000	16,908	17,823.07
		2015	14	14,000	18,800	18,000	17,586	17,619.51	17	16,000	19,400	18,547	18,547	17,425.03
		2016	12	17,500	19,000	18,123	18,245	17,929.99	15	17,000	19,500	18,700	18,568	19,762.60
		2017	16	18,700	21,900	19,949	19,897	21,349.63	17	18,200	21,200	20,103	20,103	24,719.22
		2018	14	22,000	27,200	24,825	24,735	24,271.41	18	22,000	28,500	25,208	25,215	23,327.46
		2019	NA	NA	NA	NA	NA	26,599.96	18	24,000	28,000	26,250	24,782	28,538.44
		2020	15	27,250	29,200	28,500	28,404	25,812.88	17	27,100	30,400	28,909	28,909	30,606.48
					Forecast horizon 6 months						Forecast horizon 12 months			
FAZ	SX5E	2002	14	3600	4300	4062	4023	3133.39	17	3710	4600	4300	4251	2386.41
		2003	NA	NA	NA	NA	NA	2419.51	15	2300	3200	2900	2890	2760.66
		2004	13	2500	3300	2879	2879	2811.08	14	2750	3300	3004	3008	2951.01
		2005	15	2800	3200	3050	3030	3181.54	19	3000	3350	3200	3160	3578.93
		2006	17	3350	3800	3700	3671	3648.92	18	3450	3950	3777	3754	4119.94
		2007	14	4000	4750	4208	4215	4489.77	20	3700	4600	4400	4394	4399.72
		2008	14	4200	4900	4508	4515	3352.81	18	4400	5100	4700	4726	2447.62
		2009	15	1600	3000	2500	2469	2401.69	17	1950	3350	2756	2756	2964.96
		2010	17	2400	3300	2910	2896	2573.32	20	2600	3700	3100	3124	2792.82
		2011	17	2400	3400	2950	2905	2848.53	22	2500	3350	3009	3018	2316.55
		2012	14	1700	2600	2300	2279	2264.72	22	2050	2850	2505	2510	2635.93
		2013	15	2162	2800	2626	2626	2602.59	20	2590	3050	2799	2797	3109.00
		2014	15	2750	3400	3250	3208	3228.25	23	3000	3600	3400	3344	3146.43
		2015	17	2800	3550	3300	3245	3424.30	22	3200	3720	3444	3438	3267.52
		2016	16	3145	3750	3550	3543	2864.74	22	3425	3800	3683	3665	3290.52
		2017	18	3000	3500	3271	3261	3441.88	23	3100	3500	3300	3295	3503.96
		2018	18	3450	4050	3748	3746	3395.60	23	3400	4000	3800	3793	3001.42
		2019	NA	NA	NA	NA	NA	3473.69	23	2800	3700	3300	3305	3745.16
		2020	21	3400	4000	3713	3713	3234.07	23	3500	4050	3850	3833	3552.64

HB = Handelsblatt; FAZ = Frankfurter Allgemeine Zeitung; DAX = German Stock Market Index; N = Number of forecasts issued; Min = Minimum; Max = Maximum; pts. = points; NA = not available. FAZ = Frankfurter Allgemeine Zeitung; DJI = Dow Jones Industrial Index; SX5E = Euro Stoxx 50; N = Number of forecasts issued; Min = Minimum; Max = Maximum; pts. = points; NA = not available.

Appendix B. Forecasters in the *Handelsblatt* Newspaper

1.	ABN Amro	45.	Kepler Equities
2.	Adca-Bank	46.	Kleinwort Benson Research
3.	B. Metzler Seel. Sohn & Co.	47.	LB Rheinland-Pfalz
4.	Baader Bank	48.	LBB Landesbank Berlin
5.	Baden-Württembergische Bank	49.	LBBW
6.	Bank in Liechtenstein	50.	Lehman Brothers
7.	Bank Julius Bär	51.	LGT Bank in Liechtenstein
8.	Bank of America	52.	M.M. Warburg & Co.
9.	Bank Sarasin	53.	Macquarie
10.	Bankhaus Ellwanger & Geiger	54.	Merck Finck & Co.
11.	Bankhaus Lampe	55.	Merrill Lynch
12.	Bankhaus Metzler	56.	Morgan Stanley
13.	Banque Nationale de Paris	57.	National-Bank
14.	Barclays	58.	NATIXIS
15.	Bayerische Landesbank	59.	NIBC
16.	Bayerische Vereinsbank	60.	Nomura
17.	Berenberg	61.	NordLB
18.	Bethmann Bank	62.	Oddo BHF
19.	BNP Paribas	63.	Pictet & Cie.
20.	Cheuvreux	64.	Postbank
21.	Citi	65.	Royal Bank of Scotland
22.	Commerzbank	66.	S.G. Warburg
23.	Crédit Lyonnais	67.	Sal. Oppenheim
24.	Credit Suisse	68.	Santander
25.	Daiwa Europe (Deutschland)	69.	Saxo Bank
26.	Dekabank	70.	SBC Warburg
27.	Deutsche Bank	71.	Schröder Bank
28.	Donner & Reuschel	72.	Schröder Münchmeyer Hengst
29.	Dresdner Bank	73.	Schroder Salomon Smith Barney
30.	DZ Bank	74.	Schweizerischer Bankverein
31.	Fürst Fugger Privatbank	75.	SGZ-Bank
32.	Fürstl. Castell'sche Bank	76.	Société Générale
33.	Goldman Sachs	77.	SYZ & Co.
34.	Gontard & Metallbank	78.	Targobank
35.	GZ-Bank	79.	UBS
36.	Haspa	80.	Unicredit HypoVereinsbank
37.	Hauck & Aufhäuser	81.	Union Bancaire Priveé
38.	Helaba	82.	Union Bank of Switzerland
39.	HSBC Trinkaus	83.	Vereins- und Westbank
40.	HSH Nordbank	84.	Vontobel
41.	IKB	85.	VP Bank
42.	IMI Bank	86.	Weberbank
43.	J. Safra Sarasin	87.	WestLB
44.	J.P. Morgan	88.	WGZ Bank

Appendix C. Forecasters in the *Frankfurter Allgemeine Zeitung*

1.	Adig	27.	J.P. Morgan
2.	Allianz SE	28.	Julius Bär
3.	Bankgesellschaft Berlin	29.	Landesbank Berlin
4.	Bankhaus Lampe	30.	Landesbank Rheinland-Pfalz
5.	Barclays Capital	31.	LBBW
6.	Bayern LB	32.	M.M. Warburg
7.	Berenberg	33.	Macquarie
8.	BNP Paribas	34.	Merck Finck Invest
9.	Citigroup	35.	Merrill Lynch
10.	Commerzbank	36.	Morgan Stanley
11.	CSFB	37.	Nomura
12.	Deka Bank	38.	Nord LB
13.	Deutsche Bank	39.	Oddo BHF
14.	Deutsche Bank/Postbank	40.	Postbank
15.	DIT	41.	Raiffeisen Bank International
16.	Dresdner Bank	42.	Sal. Oppenheim
17.	DWS	43.	Santander Asset Management
18.	DZ Bank	44.	Société Générale
19.	Erste Group	45.	UBS
20.	Goldman Sachs	46.	Union Bancaire Privée
21.	Helaba	47.	Union Investment
22.	HSBC Trinkaus & Burkhardt	48.	Vereins- und Westbank
23.	HSH Nordbank	49.	Weberbank
24.	HVB-Unicredit Bank	50.	WestLB
25.	IKB	51.	WGZ Bank
26.	ING Deutschland		

References

Andres, Peter, and Markus Spiwoks. 1999. Forecast Quality Matrix: A Methodological Survey of Judging Forecast Quality of Capital Market Forecasts. *Journal of Economics and Statistics* 219: 513–42.

Arthur, W. Brian, John H. Holland, Blake LeBaron, Richard Palmer, and Paul Tayler. 1997. Asset pricing under endogenous expectations in an artificial stock market. In *The Economy as an Evolving Complex System II*. Edited by W. Brian Arthur, Steven N. Durlauf and David A. Lane. Reading: Addison-Wesley.

Atmaz, Adem, Stefano Cassella, Huseyin Gulen, and Fangcheng Ruan. 2021. Contrarians, Extrapolators, and Stock Market Momentum and Reversal. Available online: https://dx.doi.org/10.2139/ssrn.3722540 (accessed on 1 November 2021).

Atsalakis, George S., and Kimon P. Valavanis. 2009. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert systems with Applications* 36: 10696–707. [CrossRef]

Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop. 2009. Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance* 28: 406–26. [CrossRef]

Baghestani, Hamid, Mohammad Arzaghi, and Ilker Kaya. 2015. On the Accuracy of Blue Chip Forecasts of Interest Rates and Country Risk Premiums. *Applied Economics* 47: 113–22. [CrossRef]

Bahrami, Afsaneh, Abul Shamsuddin, and Katherine Uylangco. 2018. Out-of-sample stock return predictability in emerging markets. *Accounting & Finance* 58: 727–50.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen. 2017. It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *Review of Financial Studies* 30: 3009–47. [CrossRef]

Benke, Holger. 2006. Was leisten Kapitalmarktprognosen?, Die Sicht eines Stiftungsmanagers. *Zeitschrift für das Gesamte Kreditwesen* 59: 902–6.

Bertella, Mario A., Felipe R. Pires, Ling Feng, and Harry Eugene Stanley. 2014. Confidence and the Stock Market: An Agent-Based Approach. *PLoS ONE* 9: e83488. [CrossRef]

Cassella, Stefano, and Huseyin Gulen. 2019. Belief-based Equity Market Sentiment. Available online: https://dx.doi.org/10.2139/ssrn.3123083 (accessed on 1 November 2021).

Cassella, Stefano, and Huseyin Gulen. 2018. Extrapolation Bias and the Predictability of Stock Returns by Price-Scaled Variables. *The Review of Financial Studies* 31: 4345–97. [CrossRef]

Chen, Shu-Heng, and Ya-Chi Huang. 2008. Risk preference, forecasting accuracy and survival dynamics: Simulations based on a multi-asset agent-based artificial stock market. *Journal of Economic Behavior & Organization* 67: 702–17.

Chen, Yi-ting, and Kendro Vincent. 2016. The Role of Momentum, Sentiment, and Economic Fundamentals in Forecasting Bear Stock Market. *Journal of Forecasting* 35: 504–27. [CrossRef]

Cowles, Alfred. 1933. Can stock market forecasters forecast? *Econometrica: Journal of the Econometric Society* 1: 309–24. [CrossRef]

De Bondt, Werner P. 1993. Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting* 9: 355–71. [CrossRef]
Diebold, Francis X., and Robert S. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13: 253–63.
Dimson, Elroy, and Paul Marsh. 1984. An analysis of brokers' and analysts' unpublished forecasts of UK stock returns. *The Journal of Finance* 39: 1257–92. [CrossRef]
Dua, Pami. 1988. Multiperiod forecasts of interest rates. *Journal of Business & Economic Statistics* 6: 381–84.
Fama, Eugene. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* 25: 383–417. [CrossRef]
Fassas, Athanasios, Stephanos Papadamou, and Dimitrios Kenourgios. 2021. Evaluating survey-based forecasts of interest rates and macroeconomic variables. *Journal of Economic Studies*. [CrossRef]
Filiz, Ibrahim, Jan René Judek, Marco Lorenz, and Markus Spiwoks. 2021. Interest rate forecasts in Latin America. *Journal of Economic Studies*. [CrossRef]
Filiz, Ibrahim, Thomas Nahmer, Markus Spiwoks, and Kilian Bizer. 2019. The accuracy of interest rate forecasts in the Asia-Pacific region: Opportunities for portfolio management. *Applied Economics* 51: 6309–32. [CrossRef]
Fraser, Patricia, and Ronald MacDonald. 1993. The efficiency of CAC stock price forecasts: A survey based perspective. *Revue Économique* 44: 991–1000.
Friend, Irwin, and Douglas Vickers. 1965. Portfolio selection and investment performance. *The Journal of Finance* 20: 391–415. [CrossRef]
Fujiwara, Ippei, Hibiki Ichiue, Yoshiyuki Nakazono, and Yosuke Shigemi. 2013. Financial markets forecasts revisited: Are they rational, stubborn or jumpy? *Economics Letters* 118: 526–30. [CrossRef]
Goyal, Amit, Ivo Welch, and Athanasse Zafirov. 2021. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II. Available online: https://dx.doi.org/10.2139/ssrn.3929119 (accessed on 1 November 2021).
Granger, Clive W., and Paul Newbold. 1974. Spurious regressions in econometrics. *Journal of Econometrics* 2: 111–20. [CrossRef]
Greenwood, Robin, and Andrei Shleifer. 2014. Expectations of Returns and Expected Returns. *The Review of Financial Studies* 27: 714–46. [CrossRef]
Guo, Hui. 2006. On the out-of-sample predictability of stock market returns. *The Journal of Business* 79: 645–70. [CrossRef]
Hein, Oliver, Michael Schwind, and Markus Spiwoks. 2012. Network Centrality and Stock Market Volatility: The Impact of Communication Topologies on Prices. *Journal of Finance and Investment Analysis* 1: 199–232.
Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia. 2020. Making Text Count: Economic Forecasting Using Newspaper Text. *Bank of England Research Paper Series* 865: 1–49. [CrossRef]
Kandasamy, Narayanan, Sarah N. Garfinkel, Lionel Page, Ben Hardy, Hugo D. Critchley, Mark Gurnell, and John M. Coates. 2016. Interoceptive Ability Predicts Survival on a London Trading Floor. *Scientific Reports* 6: 32986. [CrossRef]
Kazak, Ekaterina, and Winfried Pohlmeier. 2019. Testing out-of-sample portfolio performance. *International Journal of Forecasting* 35: 540–54. [CrossRef]
Krichene, Hazem, and Mhamed-Ali El-Aroui. 2018. Artificial stock markets with different maturity levels: Simulation of information asymmetry and herd behavior using agent-based and network models. *Journal of Economic Interaction and Coordination* 13: 511–35. [CrossRef]
Kunze, Frederik, Markus Spiwoks, Kilian Bizer, and Torsten Windels. 2018. The usefulness of oil price forecasts—Evidence from survey prediction. *Managerial and Decision Economics* 39: 427–46. [CrossRef]
Kunze, Frederik, Christoph Wegener, Kilian Bizer, and Markus Spiwoks. 2017. Forecasting European interest rates in times of financial crisis—What insights do we get from international survey forecasts? *Journal of International Financial Markets, Institutions and Money* 48: 192–205. [CrossRef]
Lakonishok, Josef. 1980. Stock market return expectations: Some general properties. *The Journal of Finance* 35: 921–31. [CrossRef]
Lofthouse, Stephen. 1996. Why Active Investment Management is Popular: The Psychology of Extraordinary Beliefs. *Journal of Interdisciplinary Economics* 7: 41–61. [CrossRef]
Mallikarjuna, Mejari, and R. Prabhakara Rao. 2019. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation* 5: 1–16. [CrossRef]
Mandelbrot, Benoit. 2004. *The (Mis)Behavior of Markets—A Fractal View of Risk, Ruin and Reward*. New York: Basic Books, pp. 186–95.
Maxwell, Michael, and Gary van Vuuren. 2019. Active investment strategies under tracking error constraints. *International Advances in Economic Research* 25: 309–22. [CrossRef]
Miah, Fazlul, Ahmed Ali Khalifa, and Shawkat Hammoudeh. 2016. Further evidence on the rationality of interest rate expectations: A comprehensive study of developed and emerging economies. *Economic Modelling* 54: 574–90. [CrossRef]
Milas, Costas, Theodore Panagiotidis, and Theologos Dergiades. 2021. Does It Matter Where You Search? Twitter versus Traditional News Media. *Journal of Money, Credit and Banking* 153: 1757–95. [CrossRef]
Mincer, Jacob A., and Victor Zarnowitz. 1969. The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectation*. Edited by Jacob Mincer. New York: Columbia University Press, pp. 3–46.
Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou. 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60: 1772–91. [CrossRef]
Nyberg, Henri. 2013. Predicting bear and bull stock markets with dynamic binary time series models. *Journal of Banking & Finance* 37: 3351–63.
Ogburn, William F. 1934. Studies in Prediction and the Distortion of Reality. *Social Forces* 13: 224–29. [CrossRef]

Oliver, Nelson, and Mehmet Pasaogullari. 2015. Interest Rate Forecasts in Conventional and Unconventional Monetary Policy Periods. *Economic commentary* 5: 1–4. [CrossRef]

Ortiz-Teran, Elena, Tomas Ortiz, Agustin Turrero, and Joaquin Lopez-Pascual. 2019. Neural implications of investment banking experience in decision-making under risk and ambiguity. *Journal of Neuroscience, Psychology, and Economics* 12: 34–44. [CrossRef]

Pierdzioch, Christian. 2015. A note on the directional accuracy of interest-rate forecasts. *Applied Economics Letters* 22: 1073–77. [CrossRef]

Ponta, Linda, and Silvano Cincotti. 2018. Traders' Networks of Interactions and Structural Properties of Financial Markets: An Agent-Based Approach. *Hindawi Complexity* 2018: 1–9. [CrossRef]

Rajab, Sharifa, and Vinod Sharma. 2019. An interpretable neuro-fuzzy approach to stock price forecasting. *Soft Computing* 23: 921–36. [CrossRef]

Ramnath, Sundaresh, Steve Rock, and Philip Shane. 2008. The Financial Analyst Forecasting Literature: A Taxonomy with Suggestions for Further Research. *International Journal of Forecasting* 24: 34–75. [CrossRef]

Spiwoks, Markus, and Oliver Hein. 2007. Die Währungs-, Anleihen- und Aktienmarktprognosen des Zentrums für Europäische Wirtschaftsforschung. *AStA Wirtschafts- und Sozialstatistisches Archiv* 1: 43–52. [CrossRef]

Spiwoks, Markus. 2004. The Usefulness of ZEW Stock Market Forecasts for Active Portfolio Management Strategies. *Journal of Economics and Statistics* 224: 557–78.

Spiwoks, Markus, Zulia Gubaydullina, and Oliver Hein. 2015. Trapped in the Here and Now - New Insights into Financial Market Analyst Behavior. *Journal of Applied Finance and Banking* 5: 29–50.

Theil, Henri. 1958. *Economic Forecasts and Policy*. Amsterdam: North Holland Publishing Company.

Theissen, Erik. 2007. An analysis of private investors' stock market return forecasts. *Applied Financial Economics* 17: 35–43. [CrossRef]

Wald, Abraham. 1943. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society* 54: 426–82. [CrossRef]

Welch, Ivo, and Amit Goyal. 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21: 1455–508. [CrossRef]

Werner, Natalie S., Katharina Jung, Stefan Duschek, and Rainer Schandry. 2009. Enhanced cardiac perception is associated with benefits in decision-making. *Psychophysiology* 46: 1123–29. [CrossRef] [PubMed]

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Yang, Xinhui, Jie Zhang, and Qing Ye. 2020. Tick size and market quality: Simulations based on agent-based artificial stock markets. *Intelligent Systems in Accounting, Finance and Management* 27: 125–41. [CrossRef]

Chapter IX

Unicorn, Yeti, Nessie, and Neoclassical Market - Legends and Empirical Evidence

Co-authored by Ibrahim Filiz, Jan René Judek, and Markus Spiwoks
Contribution Marco Lorenz: 45%

Submitted to:
German Economic Review

Published in:
Wolfsburg Working Papers, Ostfalia University of Applied Sciences, No. 22-03, Wolfsburg, August 2022. (Working Paper)

Abstract

The neoclassical market model continues to have a major influence on important economic policy decisions. In this model, the formation of equilibrium prices at the intersection of the aggregated supply and aggregated demand functions plays a central role. We examine whether the formation of equilibrium prices actually occurs. To do so, we analyze 2,217 prices for homogeneous products recorded by students in stores and online between October 2020 and May 2022. In 143 out of 146 cases, no equilibrium price emerges. The percentage price range regularly exceeds 100%. The presumed steering function of an equilibrium price does not materialize. The establishment of market mechanisms for the efficient solution of economic problems must therefore be questioned.

Keywords

Equilibrium price, price dispersion, price comparison, market equilibrium, information asymmetry, neoclassical economics.

JEL Classification

B13, D01, D11, D12, D40, D50, E13, L13, P12.

1. Introduction

Some people believe that cheerful unicorns trot through the few remaining remote primeval forests, that an intermediate form of animal and human, namely the Yeti, is at home in the heights of the Himalayas, and that an underwater monster meanders through the waters of Loch Ness in Scotland. There are eyewitness accounts for all these phenomena. To make a long story short: When critically examining all of these indications, one cannot speak of clear empirical evidence. While very few adults believe in the existence of unicorns, snow giants, and sea monsters, the idea of the neoclassical market is widespread in science and practice and finds many supporters and advocates.

Adam Smith (1776) was the first to emphasize the central importance of markets for all processes at the level of the individual economy as well as at the level of the national economy. The Austrian School propagated the market as the central element of liberal economic activity and free society, and as the decisive counter-concept to planned economy and socialism. Ludwig von Mises, Friedrich August von Hayek, Gottfried Harberler, and Oskar Morgenstern can be considered the most ardent advocates of market orientation. The influential textbook by Paul A. Samuelson (1948) contributed quite significantly to the spread of the notion of the neoclassical market. In current textbooks, such as Mankiw and Taylor (2020), it is presented unchanged: As prices rise, the willingness of (actual or potential) suppliers to provide products increases. With rising prices, however, the willingness of (actual or potential) demanders to purchase the corresponding products decreases. This results in an aggregate supply function with a positive slope and an aggregate demand function with a negative slope. The intersection of the supply and demand functions determines the equilibrium price. This equilibrium price leads to market clearing. We refer to this form of market mechanism as the neoclassical market. The neoclassical market mechanism has the merit of automatically leading to efficient outcomes (Mises, 1929 and 1940). Goods necessarily flow to the demanders who most desire or need them. Goods are necessarily provided by the suppliers who can most easily spare or produce them.

The neoclassical market model is based on a variety of assumptions (Pindyck & Rubinfeld, 2017; Varian, 2014; Samuelson & Nordhaus, 2009): 1. The traded goods are homogeneous, 2. There are many suppliers and many demanders, 3. No supplier and no demander exercises market power, 4. Any supplier and any demander can enter or leave the market at any time without costs, 5. All suppliers and all demanders are independent of each other and consider only their own situation when making decisions, i.e., there is no herd behavior and no strategic behavior, 6. There is complete information and thus no asymmetric information distribution, 7. All suppliers and all demanders are rational agents interested solely in maximizing their utility, 8. Property rights are always evident and undisputed, 9. There are no transaction costs, 10. There are no externalities.

However, it is considered a weakness of the neoclassical market model that the assumptions can often not be considered fulfilled in reality. In practical economic life, there is not a single market in which all ten model assumptions can be regarded as fulfilled (see, e.g., Ötsch, 2019; Bridel, 1997; Walker, 1993). In some markets, not even one of the assumptions is fulfilled. However, this alone cannot shake the neoclassical market model because the core statements of the model fit the everyday experience that when supply declines, prices often rise and when demand declines,

prices often fall. Even a model whose assumptions do not correspond to reality can lead to useful descriptions of the real world.

However, this does not relieve economics of the duty to empirically verify whether the neoclassical market really exists or whether it is exclusively the product of a fantasy world - similar to the case, presumably, of unicorns, Yeti, or Nessie. This leads us to the core of the problem. The neoclassical market model largely eludes empirical observation. It is plausible to assume that the aggregate supply function shows an increasing and the aggregate demand function a decreasing trend. Empirically, however, neither the one nor the other function can be represented. This will be illustrated by an example:

An innkeeper who runs the only pub in a village wants to find out how his guests would react to any price changes. So, he tries to get a picture of the aggregate demand function of his guests. For this reason, he distributes questionnaires to the guests. There they are to enter how much beer they would drink at which price level. However, his guests smell a rat and behave strategically. If the price level were higher than the current one, the guests say they would no longer come to the pub at all and would rather drink their beer at home. If the price level were lower than the current one, they make exaggerated statements about their planned consumption behavior. The innkeeper who relies on this information is in for a surprise. If he actually lowers the price, the additional consumption estimated by the survey will probably largely fail to materialize. A demand function thus eludes empirical observation.

The same applies to aggregated supply functions. For example, the water protection authority may ask the farmers in a region how much targeted floodplain area they would make available in the event of a flood, depending on how high the authority sets the compensation payments. The farmers will promise large amounts of land as compensation increases, but they do so only to drive up compensation for existing floodplains as much as possible. The strategic behavior of the providers makes it impossible to observe the actual supply function empirically.

Since neither the supply function nor the demand function can be observed, the empirical examination of the neoclassical market model has to focus on the equilibrium price. The market model assumes that all transactions of a given good at a given point in time (at least) are settled at one and the same price, namely the equilibrium price. Numerous studies have therefore collected prices for homogeneous goods from different suppliers within a market that is clearly defined in terms of space and time. This has shown again and again that prices for one and the same good can vary greatly (see Table 1).

Table 1: Synoptic literature review

Research paper	Markets covered	Methods	Conclusion
Stigler (1961)	Cars, anthracite coal	Absolute price range, standard deviation	No price equilibrium
Maynes (1976)	Life insurance, drugstore items, consumer electronics, petrol, etc.	Absolute price range	No price equilibrium
Pratt, Wise & Zeckhauser (1979)	39 different product categories, e.g., bicycles, aquariums, air conditioners	Minimum, maximum, mean, standard deviation	No price equilibrium
Dahlby & West (1986)	Car insurance premiums	Percentage price range, variance, coefficient of variation	No price equilibrium
Van Hoomissen (1988)	Refrigerators, light bulbs, books for children, various groceries	Interstore relative price variability (measure of spread based on standard deviation)	No price equilibrium
Borenstein & Rose (1994)	Flight tickets	Gini coefficient, coefficient of variation	No price equilibrium
Brynjolfsson & Smith (2000)	Books, CDs	Absolute price range, percentage price range, standard deviation	No price equilibrium
Kessner & Polborn (2000)	Life insurances	Coefficient of variation	No price equilibrium
Sorensen (2000)	Prescription drugs	Absolute price range, standard deviation, coefficient of variation, estimated margin	No price equilibrium
Brown & Goolsbee (2002)	Life insurances	Standard deviation of residuals from a regression of life insurance prices paid and various explanatory variables	No price equilibrium
Lach (2002)	Refrigerators, chicken, coffee, flour	Mean, coefficient of variation, F-test, standard deviation, time effects	No price equilibrium
Scholten & Smith (2002)	20 different retail products, e.g., groceries, toiletries, batteries, cleaning products, thermometers	Coefficient of variation	No price equilibrium
Aalto-Setälä (2003)	120 different food items	Standard deviation, mean, regression analysis	No price equilibrium
Baye, Morgan & Scholten (2004)	Consumer electronics	Minimum, absolute price range, percentage price range, coefficient of variation	No price equilibrium
Baye, Morgan & Scholten (2006)	Consumer electronics	Absolute price range, percentage price range, coefficient of variation	No price equilibrium
Hong & Shum (2006)	Books	Absolute price range	No price equilibrium
Lewis (2008)	Petrol	Standard deviation, regression analysis	No price equilibrium
Wildenbeest (2011)	14,000 products from supermarkets	Regression analysis	No price equilibrium
Vukina & Zheng (2010)	Live hogs	Minimum, maximum, mean, standard deviation, absolute price range	No price equilibrium
Moen, Wulfsberg & Aas (2020)	766 homogeneous products in 4,297 retail stores	Coefficient of variation	No price equilibrium

Empirical research on the equilibrium price begins in the second half of the 20th century, when several authors start collecting prices for homogeneous goods in shops or in magazines. Stigler (1961) discovers price dispersion in the automobile and anthracite coal markets. Maynes (1976) compares prices for identical life insurances, medicines, and nine other products and finds that different providers charge very different prices for homogeneous goods. In their influential study, Pratt, Wise & Zeckhauser (1979) extend the research to 39 different products. They show that

the formation of equilibrium prices suggested by the neo-classical theory is by no means observable in practice. Burdett & Judd (1983) counter that the empirical study of prices is not suitable for deriving convincing statements about the validity of the neoclassical market model. If the observed price dispersion can be explained by search costs, one can still speak of market equilibrium even with different prices for homogeneous goods. However, search costs cannot be quantified easily in practice, which creates a similar dilemma as with the supply and demand functions mentioned at the beginning.

From the 1980s onward, the background to the observed price deviations is increasingly investigated. In particular, factors influencing the extent of price dispersion have come into focus. It has been shown that price dispersion tends to increase when the number of suppliers is low (Baye, Morgan & Scholten, 2004; Dahlby & West, 1986) and in times of strong inflation (Van Hoomissen, 1988). Other studies find that price dispersion tends to be lower for consumer goods that are regularly repurchased (Sorensen, 2000) and for goods in mature markets (Baye, Morgan & Scholten, 2006). Moreover, deviations from the equilibrium price model are also observed within one supplier. Even for the same supplier, there can be strong price deviations for different customer groups (Borenstein & Rose, 1994) or at different times of the day (Vukina & Zheng, 2010), which also does not fit the neoclassical theory.

Technological and political events also influence the scientific debate about the equilibrium price. In the 1990s, increasing globalization and the spread of the Internet are changing the way commerce works. The transportation of raw materials and finished goods is constantly becoming faster and cheaper. New forms of communication make it possible to work efficiently with supra-regional customers and suppliers at different points in the value chain, which creates more competition. Due to all these influencing factors, the number of potential suppliers and buyers within a market increases massively. The increasing spread of the Internet also ensures more transparency for customers, who can now compare prices from different providers much more easily.

These changes lead to a revival of empirical research on equilibrium pricing in the early 2000s. However, the studies always come to the same conclusion. Globalization and the introduction of the Internet have slightly shifted the extent of price dispersion in individual market segments, but an equilibrium price in the sense of neoclassical theory is still nowhere to be observed (Brown & Goolsbee, 2002; Scholten & Smith, 2002; Brynjolfsson & Smith, 2000).

Since the 2010s, fewer and fewer price comparisons have been published to test the neoclassical market model. The new studies partly argue on the basis of data that are several decades old (e.g., Moen, Wulfsberg & Aas, 2020). With the beginning of the 2020s, however, massive political and societal cuts have again had an impact on markets. In the course of the Covid-19 pandemic, international supply chains collapsed en masse. Important products for daily use became scarce within a short period of time. Many consumers also faced changing financial conditions due to layoffs and short-time work. All of this can lead to changes in behavior on the part of both suppliers and consumers.

We see it as an important task of research to continuously test the validity of established models from theory in practice. The neoclassical market model is undoubtedly one of them. This study

therefore raises the research question of whether an equilibrium price has been established for different product categories and products in the year of the Covid-19 pandemic outbreak (2020) and the two subsequent years.

2. Data basis

To shed more light on our research question, we evaluate actual prices for homogeneous goods announced by retailers between October 2020 and May 2022. The prices were collected and documented by students of the Ostfalia University of Applied Sciences Wolfsburg in brick-and-mortar stores or online. In-store price collections were conducted in northern Germany in twelve cities (Brunswick, Einbeck, Gifhorn, Goslar, Hanover, Helmstedt, Hildesheim, Lunenburg, Peine, Salzgitter, Wolfenbüttel, and Wolfsburg). Online price collections were conducted on price comparison portals or directly on the retailers' websites. All prices were quoted in €. All price observations were documented by photos or screenshots with location and time information. The observations have been published in the series "Wolfsburg Invisible Hand Studies" (WIHSt).

The eleven studies in the WIHSt series (see Table 2) cover 146 price comparisons for 77 different products, consisting of a total of 2,217 individual price observations (see Table 3). Each price comparison thus comprises an average of 15.185 prices from different retailers for one and the same product.

Table 2: Overview of studies from the "Wolfsburg Invisible Hand Studies" series

Study	Year	Authors
WIHSt 1	2021	Kornhardt, C.
WIHSt 2	2021	Yavuz, D.
WIHSt 3	2021	Clar, F., Petrunina, J., Qitaku, A. & Zubke, L.
WIHSt 4	2021	Chmielewski, L. & Kunzmann, O.
WIHSt 5	2021	Möbius, D., Schmidt, M. & Waldhelm, S.
WIHSt 6	2021	Flemming, J., Boztepe, C. & Tawbe, S.
WIHSt 7	2022	Wenzlaff, L. & Leohold, S.
WIHSt 8	2022	Beck, O. & Ülker, S.-L.
WIHSt 9	2022	Wahlers, J. & Schulenburg, S. von der
WIHSt 10	2022	Younis, R. & Sokolowski, P.
WIHSt 11	2022	Ziegner, K. & Mützel, P.

Table 3: Number of price observations per product category

Product category	Price comparisons	Prices
Drugstore	47	634
Food	59	776
Other	40	807
Total	146	2,217

For a detailed overview per product, see Table A-1 in Appendix A.

The database is divided into 47 price comparisons for drugstore products (634 price observations), 59 price comparisons for food (776 price observations), and 40 price comparisons for other products (807 price observations). The "Other" category includes products from the areas of consumer electronics, toys, kitchen appliances, clothing, printer supplies, medicines, car accessories, and sports & outdoor. The original product designations are used below. An overview of the respective English product descriptions can be found in Table A-2 in Appendix B.

Eight products are examined in more than one Wolfsburg Invisible Hand Studies (WIHSt). These are Red Bull Classic 250ml (3×), as well as Funny-frisch "Ungarisch" 175g, Nutella Nuss-Nugat-Creme 450g, Pringles "Original" 200g, Duschdas Sport 2-in-1 250ml, Milka Alpenmilch 100g, WMF Kult X Mix & Go 0,6l, and Dr. Oetker Ristorante Salame 320g (2× each). For the remaining 69 products, prices are only examined in one WIHSt each.

The price comparisons in the studies from the WIHSt series are from 2020 (10×), 2021 (84×), and 2022 (52×). 45 price comparisons were conducted exclusively in brick-and-mortar stores, 52 price comparisons were conducted exclusively online, and 49 price comparisons were conducted both in brick-and-mortar stores and online. In the 101 price comparisons that were conducted completely or partially online, prices including shipping costs were documented in 35 cases.

3. Methods

The prices are examined with regard to the setting of an equilibrium price. An equilibrium price exists when all transactions are carried out at the same price. To this end, the lowest observed price (minimum) is compared with the highest observed price (maximum). The difference between minimum and maximum is the absolute price range. If it is different from 0, there is no equilibrium price.

In order to determine the extent of price dispersion, relative measures of dispersion are used in the form of the percentage price range and the coefficient of variation. The percentage price range is obtained by dividing the absolute price range by the minimum:

$$Percentage\ Price\ Range\ (X) = \frac{Absolute\ Price\ Range\ (X)}{Minimum\ (X)} \times 100\%$$
$$= \frac{[Maximum\ (X) - Minimum\ (X)]}{Minimum\ (X)} \times 100\%$$

A percentage price range > 100% indicates that the absolute price range is larger than the minimum. In other words, in this case the maximum is more than twice as large as the minimum, which indicates a strong deviation from an equilibrium price.

The coefficient of variation is obtained by dividing the standard deviation by the arithmetic mean of the observations:

$$Coefficient\ of\ Variation\ (X) = \frac{Standard\ Deviation\ (X)}{Arithmetic\ Mean\ (X)} = \frac{\sigma\ (X)}{\mu\ (X)}$$

A coefficient of variation > 0.1 indicates that the standard deviation is greater than 10% of the arithmetic mean, which also indicates a strong deviation from an equilibrium price.

The percentage price range and the coefficient of variation are more meaningful in that they relate the absolute price range and the standard deviation, respectively, to the price of the product. An absolute price range or a standard deviation of €1 represents a significantly more serious deviation from an equilibrium price if the observed product is priced in the order of €10 than if it is priced in the order of €1,000.

4. Results

A total of 146 price comparisons are analyzed. In 143 cases, the absolute price range, percentage price range, standard deviation, and coefficient of variation are not equal to zero. That is, in 143 of 146 cases, different prices were observed for a homogeneous good at a given time within a narrowly defined geographic area. The average percentage price range is 126.37%. The one-sample t-test supports that there is no equilibrium price ($p \leq 0.001$).

First, the price comparisons collected purely in brick-and-mortar retail stores are considered (Table 4). Here, 45 price comparisons with 398 individual price observations are carried out. The products for which a price comparison is carried out are displayed sorted according to the coefficient of variation (last column). That is, from products that show a large deviation from an equilibrium price to products that show a less large deviation from an equilibrium price.

Table 4: Comparison of prices collected in stores

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Coca-Cola Original Taste 0,33l	17	€0.59	€2.29	€0.95	€0.47	€1.70	288.14%	0.495
Nivea Deoroller Fresh pure 0%	11	€1.65	€1.95	€1.92	€0.90	€0.30	18.18%	0.469
Milka Alpenmilch 100g (WIHSt 5)	10	€0.59	€1.99	€1.03	€0.43	€1.40	237.29%	0.418
Pringles Chips Sour Cream & Onion 200g	13	€1.29	€3.89	€2.14	€0.78	€2.60	201.55%	0.364
Funny-frisch "Ungarisch" 175g (WIHSt 3)	12	€1.49	€3.29	€1.71	€0.50	€1.80	120.81%	0.294
Snickers 50g (07.08.21)	8	€0.69	€1.39	€0.86	€0.25	€0.70	101.45%	0.291
Snickers 50g (30.08.21)	8	€0.69	€1.39	€0.86	€0.25	€0.70	101.45%	0.291
Pflaster Hansaplast "Classic"	20	€2.95	€5.55	€3.75	€1.06	€2.60	88.14%	0.282
Konfitüre Schwartau Extra Erdbeere 340g	7	€1.39	€3.49	€2.32	€0.62	€2.10	151.08%	0.266
Red Bull Classic 250ml (WIHSt 11)	16	€1.05	€2.55	€1.55	€0.40	€1.50	142.86%	0.258
Honig Langnese "Flotte Biene" 250g	5	€2.99	€4.49	€3.29	€0.67	€1.50	50.17%	0.204
Jägermeister 0,7l	9	€10.49	€18.99	€12.77	€2.43	€8.50	81.03%	0.190
UHU Kleber 21g (09.08.21)	7	€1.89	€2.99	€2.15	€0.39	€1.10	58.20%	0.181
UHU Kleber 21g (30.08.21)	6	€1.89	€2.99	€2.18	€0.39	€1.10	58.20%	0.179
Dr. Oetker Ristorante Salame 320g (WIHSt 5)	8	€1.59	€2.49	€2.28	€0.39	€0.90	56.60%	0.171
Red Bull Sugarfree 250ml	19	€0.87	€1.59	€1.18	€0.20	€0.72	82.76%	0.170
Nutella Nuss-Nugat-Creme 450g (WIHSt 4)	7	€2.99	€4.29	€3.18	€0.49	€1.30	43.48%	0.154

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Zahnpasta Elmex "Kariesschutz"	24	€2.55	€4.20	€3.19	€0.47	€1.65	64.71%	0.146
Pril Kraftgel Ultra Plus	19	€0.99	€1.45	€1.33	€0.19	€0.46	46.46%	0.143
Pringles "Original" 200g (WIHSt 5)	9	€2.59	€3.69	€2.71	€0.37	€1.10	42.47%	0.137
Duschgel Kneipp "Lebensfreude"	12	€2.95	€4.45	€3.11	€0.42	€1.50	50.85%	0.136
Niemand Dry Gin 0,5l (30.08.21)	3	€28.99	€36.99	€34.31	€4.61	€8.00	27.60%	0.134
UNO Standard	9	€6.99	€9.99	€8.77	€1.11	€3.00	42.92%	0.127
Converse Chuck Taylor All Star High	7	€49.99	€75.00	€71.41	€8.75	€25.01	50.03%	0.123
Géramont "Classic" 200g	6	€1.66	€2.44	€2.29	€0.28	€0.78	46.99%	0.122
Niemand Dry Gin 0,5l (06.08.21)	4	€28.99	€36.99	€34.98	€3.99	€8.00	27.60%	0.114
Maggi Würze 250g	20	€1.25	€2.39	€1.80	€0.19	€1.14	91.20%	0.103
Nivea Soft 200ml	13	€2.85	€3.49	€2.96	€0.22	€0.64	22.46%	0.074
WMF Kult X Mix & Go 0,6l (WIHSt 11)	3	€29.99	€34.99	€33.31	€2.35	€5.00	16.67%	0.071
Xiaomi Scooter 1S	3	€395.99	€449.00	€413.66	€24.99	€53.01	13.39%	0.060
Leibniz Keks'N Crem Choco 228g	9	€1.69	€1.99	€1.85	€0.11	€0.30	17.75%	0.058
AirPods 2. Gen. / MV7N2ZM/A (30.08.21)	4	€132.99	€149.00	€134.32	€7.58	€16.01	12.04%	0.056
TomTom "Go Discover 7"	3	€269.00	€299.00	€281.33	€12.82	€30.00	11.15%	0.046
AirPods 2. Gen. / MV7N2ZM/A (18.08.21)	4	€126.95	€136.99	€129.97	€4.78	€10.04	7.91%	0.037
Schauma 7 Kräuter Shampoo	17	€1.65	€1.99	€1.94	€0.07	€0.34	20.61%	0.036
HP 302 Cyan/Magenta/Gelb Druckerpatrone	5	€22.90	€24.99	€23.37	€0.81	€2.09	9.13%	0.035
Hipp "Ultra Sensitiv" Feuchttücher 4er Pack	4	€3.45	€3.79	€3.62	€0.12	€0.34	9.86%	0.034
Hipp "Zart Pflegend" Feuchttücher 4er Pack	4	€3.45	€3.65	€3.54	€0.09	€0.20	5.80%	0.025
Barilla Penne Rigate 500g	6	€1.65	€1.69	€1.68	€0.02	€0.04	2.42%	0.012
I Love Extreme Mascara "Volume"	3	€2.75	€2.79	€2.76	€0.02	€0.04	1.45%	0.007
Pampers "Premium Protection" 26 Stück	6	€3.95	€3.99	€3.96	€0.02	€0.04	1.01%	0.005
Aptamil "Pronatura PRE" 800g	5	€15.95	€15.99	€15.97	€0.02	€0.04	0.25%	0.001
Duschdas Sport 2-in-1 250ml (WIHSt 5)	6	€1.25	€1.25	€1.25	€0.00	€0.00	0.00%	0.000
Pampers "Baby Dry" 21 Stück	3	€2.95	€2.95	€2.95	€0.00	€0.00	0.00%	0.000
Bebe "Creme Intensivpflege" 50ml	4	€3.45	€3.45	€3.45	€0.00	€0.00	0.00%	0.000
Total	398							0.145

= Number of observations; Min = Minimum; Max = Maximum; x̄ = Arithmetic mean; σ = Standard deviation; APR = Absolute price range, PPR = Percentage price range; CV = Coefficient of variation.

For the product “Coca-Cola Original Taste 0,33l", 17 individual price observations are collected from 17 different retailers. While the observed minimum sales price is €0.59, the observed maximum price for this product is €2.29. The mean of the price observations is €0.95, and the standard deviation is €0.47. The absolute price range of €1.70 is obtained by subtracting the minimum from the maximum. The percentage price range is 288.14%, showing that the absolute price range is almost three times as high as the minimum price. The coefficient of variation of 0.495 also shows that there is a strong deviation from an equilibrium price.

It is remarkable that for the three goods in the last three rows of Table 4, for which all observed prices are identical (coefficient of variation = 0.000), only very few prices are recorded (n = 3; n = 4; n = 6). One of these products is "Duschdas Sport 2-in-1, 250ml" in WIHSt 5. The same product is also observed in WIHSt 3, but there at 29 different retailers in stores and online. WIHSt 3 determines an absolute price range of €0.50 for this product and therefore no equilibrium price.

Overall, the prices collected in brick-and-mortar stores yield an average coefficient of variation of 0.145. The majority of price comparisons in brick-and-mortar stores show that an equilibrium price cannot be observed.

Table 5 lists the price comparisons carried out purely in online retailing. Shipping costs are not taken into account at first. Here, 28 price comparisons with 497 individual price observations are carried out. For the product "Bebe Creme Intensivpflege 50ml", the mean of €4.67 and the standard deviation of €2.26 result in the highest coefficient of variation of 0.485 and thus the greatest deviation from an equilibrium price.

The largest percentage price range is observed for the product "UHU Kleber 21g" at 602.78%. That is, the most expensive retailer offers the product at a selling price more than six times higher than the least expensive retailer. Overall, the high percentage price ranges in online retailing show that there are large price differences among suppliers on the Internet. This is also reflected in the higher average coefficient of variation of 0.219 in online retailing compared with stationary retailing. In the price comparisons carried out among online retailers without taking shipping costs into account, an equilibrium price cannot be observed.

Table 5: Comparison of prices collected online excluding shipping costs

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Bebe "Creme Intensivpflege" 50ml	6	€3.44	€9.70	€4.67	€2.26	€6.26	181.98%	0.485
Nivea Soft 200ml	8	€2.66	€8.02	€3.82	€1.81	€5.36	201.50%	0.474
UHU Kleber 21g (09.08.21)	43	€1.08	€7.59	€2.35	€1.11	€6.51	602.78%	0.472
UHU Kleber 21g (30.08.21)	51	€1.08	€7.59	€2.34	€1.07	€6.51	602.78%	0.457
Zahnpasta Elmex "Kariesschutz"	9	€2.89	€7.70	€3.82	€1.58	€4.81	166.44%	0.414
WMF Kult X Mix & Go 0,6l (WIHSt 5)	6	€29.99	€64.99	€40.16	€13.86	€35.00	116.71%	0.345
Duschgel Kneipp "Lebensfreude"	10	€2.39	€5.99	€3.31	€1.06	€3.60	150.63%	0.321
Nivea Deoroller Fresh pure 0%	7	€1.65	€3.73	€2.32	€0.74	€2.08	126.06%	0.319
Red Bull Classic 250ml (WIHSt 11)	4	€1.20	€2.44	€1.70	€0.46	€1.24	103.33%	0.271
Pampers "Premium Protection" 26 Stück	9	€3.82	€8.18	€5.45	€1.45	€4.36	114.14%	0.267
WMF Kult X Mix & Go 0,6l (WIHSt 11)	16	€29.99	€59.99	€37.04	€9.35	€30.00	100.03%	0.252
Honig Langnese "Flotte Biene" 250g	7	€2.72	€5.10	€3.76	€0.85	€2.38	87.50%	0.226
AirPods 2. Gen. / MV7N2ZM/A (18.08.21)	76	€123.80	€345.69	€150.94	€30.97	€221.89	179.23%	0.205
Pflaster Hansaplast "Classic"	12	€2.80	€4.79	€3.58	€0.60	€1.99	71.07%	0.169
UNO Standard	11	€6.52	€10.71	€8.21	€1.37	€4.19	64.26%	0.167
Konfitüre Schwartau Extra Erdbeere 340g	7	€2.09	€3.29	€2.63	€0.44	€1.20	57.42%	0.166
AirPods 2. Gen. / MV7N2ZM/A (30.08.21)	89	€119.95	€236.29	€150.63	€23.29	€116.34	96.99%	0.155
Nutella Nuss-Nugat-Creme 450g (WIHSt 4)	7	€2.84	€4.15	€3.39	€0.50	€1.31	46.13%	0.148
HP 302 Cyan/Magenta/Gelb Druckerpatrone	15	€17.64	€27.99	€22.48	€3.06	€10.35	58.67%	0.136
Niemand Dry Gin 0,5l (30.08.21)	24	€26.01	€39.90	€32.33	€3.57	€13.89	53.40%	0.110
Niemand Dry Gin 0,5l (06.08.21)	23	€26.01	€39.90	€32.54	€3.59	€13.89	53.40%	0.110
Converse Chuck Taylor All Star High	11	€55.95	€79.00	€72.10	€6.06	€23.05	41.20%	0.084
TomTom "Go Discover 7"	11	€214.46	€299.99	€271.66	€22.29	€85.53	39.88%	0.082
Algemarina Trockenshampoo 200ml	5	€2.95	€3.49	€3.27	€0.26	€0.54	18.31%	0.081
Head&Shoulders Apple Fresh 300ml	13	€3.29	€3.99	€3.90	€0.26	€0.70	21.28%	0.067
Sony Playstation 5 Disc Version	5	€795.99	€944.00	€887.72	€48.94	€148.01	18.59%	0.055
Xiaomi Scooter 1S	6	€395.00	€499.00	€415.58	€18.56	€104.00	26.33%	0.045
I Love Extreme Mascara "Volume"	6	€2.75	€3.09	€2.83	€0.12	€0.34	12.36%	0.042
Total	497							0.219

= Number of observations; Min = Minimum; Max = Maximum; x̄ = Arithmetic mean; σ = Standard deviation; APR = Absolute price range, PPR = Percentage price range; CV = Coefficient of variation.

A similar picture emerges when looking at price comparisons for online retailing including shipping costs (Table 6). Here 24 price comparisons with 459 individual price observations are considered. If shipping costs are incurred, the inclusion of the shipping costs in the sales prices may increase the maximum price by the amount of the shipping costs. If no shipping costs are incurred or if these are included in the product price, the sales price does not increase, and the minimum price may remain constant. This can affect the average price, the standard deviation, the absolute and percentage price range, and the coefficient of variation.

Both the highest coefficient of variation (0.515) and the highest percentage price range (685.83%) are found for the product "Red Bull Classic 250ml." Overall, across the 24 price comparisons, the average coefficient of variation for online retailing including shipping costs is 0.200. This value is slightly lower than for online retailing excluding shipping costs. Likewise, an equilibrium price cannot be observed for online retailers including shipping costs in any price comparison.

Table 6: Comparison of prices collected online including shipping costs

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Red Bull Classic 250ml (WIHSt 11)	4	€1.20	€9.43	€5.92	€3.05	€8.23	685.83%	0.515
Nivea Deoroller Fresh pure 0%	7	€1.65	€9.14	€5.51	€2.18	€7.49	453.94%	0.396
WMF Kult X Mix & Go 0,6l (WIHSt 5)	6	€29.99	€68.94	€42.96	€15.08	€38.95	129.88%	0.351
Duschgel Kneipp "Lebensfreude"	10	€2.39	€11.98	€7.08	€2.37	€9.59	401.26%	0.335
Konfitüre Schwartau Extra Erdbeere 340g	7	€2.09	€9.28	€7.18	€2.39	€7.19	344.02%	0.333
WMF Kult X Mix & Go 0,6l (WIHSt 11)	16	€29.99	€63.94	€40.81	€9.60	€33.95	113.20%	0.235
Nivea Soft 200ml	8	€6.16	€10.92	€8.13	€1.73	€4.76	77.27%	0.213
Honig Langnese "Flotte Biene" 250g	7	€6.96	€12.09	€8.79	€1.81	€5.13	73.71%	0.206
AirPods 2. Gen. / MV7N2ZM/A (18.08.21)	76	€124.55	€345.69	€153.67	€30.53	€221.14	177.55%	0.199
Pflaster Hansaplast "Classic"	12	€2.95	€7.94	€6.92	€1.36	€4.99	169.15%	0.196
Nutella Nuss-Nugat-Creme 450g (WIHSt 4)	7	€5.98	€9.95	€8.27	€1.40	€3.97	66.39%	0.170
Pampers "Premium Protection" 26 Stück	9	€6.81	€12.96	€9.94	€1.65	€6.15	90.31%	0.166
UHU Kleber 21g (09.08.21)	43	€3.39	€9.86	€7.56	€1.21	€6.47	190.86%	0.160
UHU Kleber 21g (30.08.21)	51	€3.39	€10.45	€7.61	€1.17	€7.06	208.26%	0.154
AirPods 2. Gen. / MV7N2ZM/A (30.08.21)	89	€119.95	€236.29	€153.42	€22.95	€116.34	96.99%	0.150
Zahnpasta Elmex "Kariesschutz"	9	€5.88	€8.70	€7.09	€1.01	€2.82	47.96%	0.143
UNO Standard	11	€6.99	€13.68	€11.00	€1.54	€6.69	95.71%	0.140
I Love Extreme Mascara "Volume"	6	€6.25	€8.69	€7.04	€0.90	€2.44	39.04%	0.129
Bebe "Creme Intensivpflege" 50ml	4	€7.13	€9.70	€8.04	€1.00	€2.57	36.04%	0.124
HP 302 Cyan/Magenta/Gelb Druckerpatrone	15	€20.63	€30.87	€25.67	€2.99	€10.24	49.64%	0.117
Niemand Dry Gin 0,5l (30.08.21)	24	€31.24	€45.80	€37.79	€3.98	€14.56	46.61%	0.105
Niemand Dry Gin 0,5l (06.08.21)	23	€31.05	€43.98	€38.01	€3.67	€12.93	41.64%	0.097
Converse Chuck Taylor All Star High	11	€58.85	€82.94	€73.89	€6.36	€24.09	40.93%	0.086
Algemarina Trockenshampoo 200ml	4	€6.48	€7.99	€7.20	€0.57	€1.51	23.30%	0.079
Total	459							0.200

= Number of observations; Min = Minimum; Max = Maximum; x̄ = Arithmetic mean; σ = Standard deviation; APR = Absolute price range, PPR = Percentage price range; CV = Coefficient of variation.

The ability to ship products enables an online retailer to operate in the same geographical area as a brick-and-mortar retailer. In this way, the number of retailers offering a product in a narrowly defined geographic area can increase significantly. Therefore, it is also necessary to consider online retail and brick-and-mortar retail together. In Table 7, the aggregated results of stationary

and online trade are presented excluding shipping costs. Here, 38 price comparisons with 692 individual price observations are considered.

Table 7: Comparison of prices collected in stores and online excluding shipping costs

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Milka "Haselnussschokolade“ 100g	6	€0.57	€1.89	€1.11	€0.53	€1.32	231.58%	0.478
Red Bull Classic 250ml (WIHSt 1)	14	€0.85	€2.79	€1.52	€0.68	€1.94	228.24%	0.447
Funny-frisch "Ungarisch" 175g (WIHSt 1)	12	€1.34	€3.59	€1.68	€0.66	€2.25	167.91%	0.393
Red Bull Classic 250ml (WIHSt 3)[1]	15	€0.88	€2.75	€1.40	€0.54	€1.87	212.50%	0.384
Extra Professional White Kaugummi 50 Stück	12	€2.25	€4.50	€3.02	€1.03	€2.25	100.00%	0.341
Nivea Dry Impact Deo 150ml	7	€1.75	€3.53	€2.07	€0.65	€1.78	101.71%	0.314
Pringles "Original" 200g (WIHSt 1)	12	€1.15	€4.00	€2.51	€0.74	€2.85	247.83%	0.295
Haribo Happy Cola 200g	10	€0.69	€1.79	€1.19	€0.35	€1.10	159.42%	0.291
Odol-med3 Zahnpasta Extra White 125ml	20	€0.99	€2.99	€1.54	€0.43	€2.00	202.02%	0.279
Heineken Pils 6 x 0,33l	10	€4.85	€10.26	€6.00	€1.67	€5.41	111.55%	0.278
Toffifee 125g	29	€1.35	€2.89	€1.60	€0.45	€1.54	114.07%	0.278
Aspirin 500mg (20 Tabletten)	9	€3.80	€7.49	€5.71	€1.50	€3.69	97.11%	0.263
Airwaves Kaugummis Cool Cassis 12 Stück	22	€0.69	€1.59	€1.02	€0.26	€0.90	130.43%	0.255
Baby Einstein Magic Touch Piano	32	€23.99	€69.99	€33.39	€8.39	€46.00	191.75%	0.251
WMF Toaster Stelio Edelstahl	15	€36.85	€69.99	€46.88	€10.95	€33.14	89.93%	0.234
Pom-Bär Original 75g	32	€0.79	€2.59	€1.22	€0.28	€1.80	227.85%	0.226
Milka Alpenmilch 100g (WIHSt 8)	20	€0.55	€1.70	€1.11	€0.22	€1.15	209.09%	0.198
Haribo Goldbären 200g	20	€0.65	€1.18	€0.87	€0.17	€0.53	81.54%	0.195
Tempo Taschentücher 30 x 10 Stück	22	€2.85	€4.51	€3.17	€0.57	€1.66	58.25%	0.180
Uncle Ben’s Express Langkornreis 250g	18	€1.29	€2.49	€1.76	€0.31	€1.20	93.02%	0.178
JBL Flip 5	30	€84.90	€156.25	€103.43	€17.73	€71.35	84.04%	0.171
Ritter Sport Voll-Nuss 100g	28	€1.36	€2.69	€1.46	€0.25	€1.33	97.79%	0.168
Ritter Sport Alpenmilch 100g	20	€0.69	€1.39	€1.19	€0.18	€0.70	101.45%	0.151
Milka Luflée Schokolade 100g	9	€0.69	€1.29	€1.13	€0.17	€0.60	86.96%	0.148
Airwaves Strong Kaugummi 12 Stück	9	€0.75	€1.10	€0.92	€0.13	€0.35	46.67%	0.145
Pantene PRO-V Repair & Care 300ml	10	€1.99	€2.99	€2.67	€0.36	€1.00	50.25%	0.136
Dr. Oetker Ristorante Salame 320g (WIHSt 9)	17	€1.79	€2.99	€2.75	€0.36	€1.20	67.04%	0.133
Big Bobby Car Classic Sansibar	9	€45.85	€64.90	€55.12	€6.87	€19.05	41.55%	0.125
Lindt Lindor Kugel Milch 100g	11	€2.04	€2.99	€2.75	€0.31	€0.95	46.57%	0.113
FIFA 21 (Playstation 4)	39	€49.95	€80.20	€61.58	€6.85	€30.25	60.56%	0.111
Maggi Ravioli in Tomatensauce 800g	15	€1.39	€2.49	€2.00	€0.22	€1.10	79.14%	0.108
PS4 Wireless Dualshock Controller, V2	43	€48.98	€76.31	€58.72	€5.21	€27.33	55.80%	0.089
FIFA 22 (Playstation 5)	11	€59.55	€79.99	€71.90	€6.28	€20.44	34.32%	0.087
Duschdas Duschgel Sport 2-in-1 250ml	29	€1.25	€1.75	€1.29	€0.11	€0.50	40.00%	0.084
Nintendo Switch	37	€306.87	€421.28	€335.88	€26.25	€114.41	37.28%	0.078
Head&Shoulders Classic Clean 300ml	20	€3.50	€4.29	€3.91	€0.17	€0.79	22.57%	0.044
KTM Radical Kids Training Bike	7	€116.00	€129.99	€118.07	€4.92	€13.99	12.06%	0.042
Nutella Nuss-Nugat-Creme 450g	11	€2.99	€3.29	€3.06	€0.11	€0.30	10.03%	0.036
Total	692							0.203

= Number of observations; Min = Minimum; Max = Maximum; x̄ = Arithmetic mean; σ = Standard deviation; APR = Absolute price range, PPR = Percentage price range; CV = Coefficient of variation.

[1] Here, the special offer price also documented in the study is noted.

The highest coefficient of variation is 0.478, for the product "Milka Haselnussschokolade 100g." The cheapest retailer offers the product for €0.57 and the most expensive retailer for €1.89. This results in an absolute price range of €1.32, which is more than double the price of the cheapest vendor. This is also shown by the percentage price range of 231.58%.

For the joint consideration of stationary trade and online trade across all 38 price comparisons, the average coefficient of variation is 0.203. The inclusion of online retailers in addition to stationary retailers thus leads to an overall increase in the coefficient of variation. The extent of price dispersion is similar to that of pure online retailing. An equilibrium price cannot be determined here either.

Finally, the brick-and-mortar retailers and the online retailers are also analyzed jointly, taking into account the shipping costs incurred to transport the goods to the corresponding place (Table 8). Here, 11 price comparisons are documented with 171 individual price observations. At brick-and-mortar retailers, a good can be purchased directly at the called retail price. Online retailers may charge additional shipping costs to transport the goods to the end customer. As a result, the maximum selling price at online retailers increases due to the inclusion of shipping costs, while the selling price at brick-and-mortar retailers remains constant.

This is particularly noticeable when determining the coefficient of variation for the product "Milka Alpenmilch 100g." Here, the standard deviation of €2.68 is greater than the arithmetic mean of €2.53, resulting in a coefficient of variation of 1.059. The maximum price here is more than 18 times higher than the selling price of the cheapest supplier. The percentage price range of 1,721.82% also impressively shows that there is a particularly strong deviation from an equilibrium price in this price comparison.

Table 8: Comparison of prices collected in stores and online including shipping costs

Product	#	Min	Max	x̄	σ	APR	PPR	CV
Milka Alpenmilch 100g (WIHSt 8)	20	€0.55	€10.02	€2.53	€2.68	€9.47	1,721.82%	1.059
Ritter Sport Alpenmilch 100g	20	€0.69	€6.19	€1.88	€1.63	€5.50	797.10%	0.866
Haribo Goldbären 200g	20	€0.65	€8.13	€3.50	€2.90	€7.48	1,150.77%	0.829
Uncle Ben's Express Langkornreis 250g	18	€1.29	€9.48	€2.92	€2.34	€8.19	634.88%	0.800
Head&Shoulders Classic Clean 300ml	20	€3.50	€10.49	€5.20	€2.23	€6.99	199.71%	0.429
Baby Einstein Magic Touch Piano	32	€26.98	€69.99	€37.31	€8.04	€43.01	159.41%	0.216
Big Bobby Car Classic Sansibar	9	€49.75	€64.90	€56.22	€5.96	€15.15	30.45%	0.106
Aspirin 500mg (20 Tabletten)	9	€6.45	€8.74	€7.73	€0.70	€2.29	35.50%	0.091
Vilsa Classic 12 x 0,7l Kasten[2]	6	€4.92	€5.99	€5.30	€0.38	€1.07	21.75%	0.072
KTM Radical Kids Training Bike	6	€116.08	€135.94	€120.24	€7.43	€19.86	17.11%	0.062
Toniebox Starterset inkl. Kreativtonie	11	€71.76	€83.43	€78.01	€3.29	€11.67	16.26%	0.042
Total	171							0.416

= Number of observations; Min = Minimum; Max = Maximum; x̄ = Arithmetic mean; σ = Standard deviation; APR = Absolute price range, PPR = Percentage price range; CV = Coefficient of variation.

[2] The minimum order value is higher than the observed price for one retailer.

Overall, the joint analysis of brick-and-mortar retailers and online retailers including shipping costs shows a significantly higher average coefficient of variation of 0.416. An equilibrium price cannot be observed in the aggregated analysis of brick-and-mortar retailers and online retailers including shipping costs either.

Looking at the results as a whole, it becomes clear that the price comparisons in brick-and-mortar retailing with a mean coefficient of variation of 0.145 are clearly far from an equilibrium price. The setting of equilibrium prices is not observed at all, with the exception of three price comparisons with a low number of price observations in each case.

Looking at online retail, the average coefficient of variation is 0.219 (without taking shipping costs into account) and 0.200 (with taking shipping costs into account). Here it is already clear that the prices are more widely spread and that there is a greater deviation from an equilibrium price online than in stationary retail.

The joint analysis of brick-and-mortar retailers and online retailers shows an average coefficient of variation of 0.203 (without taking shipping costs into account) and 0.416 (with taking shipping costs into account). The mean coefficient of variation with shipping costs taken into account shows a significant deviation from an equilibrium price. This may be due to the fact that the shipping costs charged by an online retailer result in a higher maximum price, while the sales price remains constant for stationary retailers. This results in a higher overall price spread between brick-and-mortar retailers and online retailers, which leads to a greater deviation from an equilibrium price.

5. Discussion

Empirical results of the past 60 years clearly show that equilibrium prices do not occur, even for absolutely homogeneous goods. This is true even if one sets narrow geographic boundaries (a city) and considers only short time periods (< 1 day).

The neoclassical market model can be decomposed into three main components: 1. Aggregate supply function, 2. Aggregate demand function, 3. Equilibrium price. As already stated in the introduction, the first two main components escape empirical observation. This is because both suppliers and demanders have clearly discernible motives for never giving honest, but always strategically distorted information about their willingness to supply or demand at different price levels.

Therefore, the focus of empirical research on the neoclassical market model must be on observing price differences. The research results presented here support the findings of numerous previous studies. With very few exceptions, no equilibrium price emerges.

One might have expected that the increased importance of the Internet would contribute to a reduction in search costs and information costs, so that the empirically observable price deviations would decrease significantly. However, the price observations presented here do not indicate this at all.

Contrary to what the neoclassical market model suggests, we are dealing with a highly fragmented market. Even at one place and at one time, transactions of a homogeneous good are carried out at quite different prices. Aggregation of demand and aggregation of supply do not occur in reality. The demanders do not act as a group. Instead, with their respective demands they are fragmented into small groups or even completely isolated from the other demanders. The situation is no different for suppliers. They, too, do not act as a group. Their offerings are also fragmented or even isolated. No retailer has an overview of the entire demand. No customer has an overview of all the suppliers. Thus, some suppliers always meet some demanders without knowing or paying attention to the overall situation of aggregated supply and aggregated demand. Individual demands meet individual offers in a completely unconnected and uncoordinated way. This leads to transactions that show entirely different price levels - even for completely homogeneous goods, even at one and the same place and at one and the same time.

The reasons for this structural market fragmentation lie primarily in the non-fulfillment of the model assumptions of the neoclassical market model. In economic reality, there are herd behavior, strategic behavior, asymmetric information distributions, externalities, search costs, information costs, negotiation costs, decision costs, monitoring costs, and enforcement costs. Real economic agents are, as a rule, far from rationally seeking their pecuniary utility maximum. On the contrary, many people tend to behave irrationally at least occasionally. They accept higher prices because a store offers nice parking spaces, because they like the shopkeeper or because they are happy to chat with the nice staff from time to time.

The neoclassical market model is unquestionably the best-known and most influential model that economic science has ever produced. Even economic laymen know the representation of the aggregate demand function, the aggregate supply function, and the formation of an equilibrium price. But what is left of this model if two main components are not empirically observable and empirical observations of the third main component regularly point to the conclusion that market activity is inaccurately described by the neoclassical market model? A sober assessment must lead to the conclusion that the neoclassical market model consists to one half of wishful thinking and to the other half of (more or less esoteric) beliefs. The neoclassical market model thus appears to belong to the same category as the unicorn, the Yeti, and the Loch Ness monster.

This sobering finding makes two consequences inevitable:

1. Economic theory must produce a new market model that adequately reflects the fragmentation of markets. So far, economic research has been too comfortable. It is often conceded that the neoclassical market model does not (quite) accurately describe reality. But in the same breath, the view is often expressed that, on the whole, things will probably work out more or less as in this model. This attitude, however, is unworthy of a science. The equilibrium price is said to have a steering function. This steering function is connected with an efficiency expectation. If, however, no equilibrium price is achieved, the assumed steering function of the price does not occur, and the efficiency promise remains unfulfilled. Real markets thus deviate fundamentally from the neoclassical market model. For this reason, it is wrong to consider market orientation as the solution to almost all economic problems, as the protagonists of the Austrian School have done in the past 100 years.

2. Economic policy must no longer follow the efficiency promise of market orientation. Generations of economic policymakers believed that the establishment of market mechanisms would automatically lead to efficient outcomes. Public health policy in Germany can be seen as an example of the failure of this approach. For about 40 years, every Minister of Health has been given the task of ensuring more competition in the healthcare system and thus contributing to cost containment. In the course of this market orientation, many hospitals were privatized and competition between hospitals was stimulated. It was hoped that this would uncover hidden reserves of personnel and materials and lead to more efficient and cost-effective healthcare. The result, however, was that high-cost healthcare services in particular were performed more frequently and that costs in the hospital system continued to rise unchecked. In 2019, 315 artificial hip joints were implanted per 100,000 inhabitants in Germany. That is almost twice as many interventions as the average for OECD countries, where only 174 corresponding operations per 100,000 inhabitants occurred in 2019 (OECD, 2021). In a sector as strongly characterized by asymmetric information distributions as the healthcare sector, competition cannot lead to efficient market outcomes. A patient cannot independently judge whether an artificial hip joint is the appropriate treatment. A patient must rely on the judgment of a physician. If, however, the latter is encouraged to perform as many hip operations as possible by means of correspondingly high remuneration, competition between hospitals will not lead to greater efficiency, but rather to increasing misuse in public healthcare.

In view of the empirical results presented in this study, we should address these challenges in economic theory and economic policy with great commitment.

6. Summary

The neoclassical market model enjoys great popularity and continuous dissemination in academic teaching. In the neoclassical market model, rising prices mean that suppliers are more willing to provide goods. At the same time, however, rising prices reduce the willingness of demanders to purchase these goods. The resulting aggregate supply and demand function form an intersection which characterizes the equilibrium price and, according to the theory, leads to market clearing. The neoclassical market model, however, has a weakness in its model assumptions, which often cannot be regarded as fulfilled in reality.

This study aims to empirically test the validity of the neoclassical market model and to determine whether a neoclassical market can be observed in reality. For this purpose, price observations of homogeneous goods within a narrow geographical area at a specific point in time are conducted and analyzed. According to the neoclassical market model, a homogeneous good should have an equilibrium price and be traded at the same price by different sellers within a spatially and temporally delimited market. In academic discourse, similar price observations have repeatedly revealed widely varying prices for homogeneous goods (see, for example, Vukina & Zheng, 2010; Brynjolfsson & Smith, 2000; Borenstein & Rose, 1994; Dahlby & West, 1986; Pratt, Wise & Zeckhauser, 1979). Not only the massive political and societal influences on markets with the beginning of the 2020s (Covid-19 pandemic, collapse of international supply chains, war in

Ukraine, etc.), but also the progressive development of online trade mean that the validity of empirical findings has to be permanently verified by science.

Between October 2020 and May 2022, students at Ostfalia University of Applied Sciences in Wolfsburg conduct 146 price comparisons for 77 different goods with a total of 2,217 individual price observations. They record 59 price comparisons for food items, 47 price comparisons for drugstore items, and 40 price comparisons for other products. The price comparisons take place both in brick-and-mortar retail and in online retail. We analyze the recorded prices both separately for each type of retail and aggregated, and also consider the impact of any shipping costs that may apply. We consider the percentage price range of the observed goods and determine the coefficients of variation to analyze the extent of deviation from an equilibrium price.

It turns out that in 143 out of 146 price comparisons, the percentage price range, the standard deviation, and the coefficient of variation are non-zero. The other three price comparisons are based on very few observations. In another study analyzed, a second price comparison was carried out for one of these three goods with a significantly higher number of individual price observations, and it was found that an equilibrium price cannot be observed. According to our data, the setting of an equilibrium price for a homogeneous good cannot be observed in a spatially and temporally delimited market (p-value of one-sample t-test ≤ 0.001).

The strongest deviation from an equilibrium price is found in the aggregated analysis of brick-and-mortar and online retail including shipping costs with a coefficient of variation of 0.416. In brick-and-mortar retailing only, we find the smallest deviation from an equilibrium price with a coefficient of variation of 0.145. The fact that online retailers can operate in the same geographic area as brick-and-mortar retailers seems to result in a stronger deviation from an equilibrium price overall.

Our results are consistent with previous academic findings in the literature. Despite recent massive political and social influences on markets, our results support previous empirical findings that do not observe an equilibrium price according to the neoclassical market model in reality. Transactions of homogeneous goods are carried out at different prices. In contrast to the neoclassical market model, market activity in reality is highly fragmented. There is no aggregation of supply and demand. Suppliers do not act as one group, within suppliers there are many groups that act separately. Consumers do not act as one group either, within consumers there are also many groups that act separately from each other. The entirety of supply and demand cannot be processed by individual actors. As a result, transactions occur at different price levels, even though the goods in question are homogeneous.

Economic theory must take the fragmentation of markets adequately into account and produce a new market model. Economic policy should abandon the naïve notion that the establishment of market mechanisms alone will produce efficient results.

References

Aalto-Setälä, V. (2003). Explaining Price Dispersion for Homogeneous Grocery Products, *Journal of Agricultural & Food Industrial Organization*, *1*(1), 1-16.

Baye, M. R., Morgan, J., & Scholten, P. (2006). Persistent Price Dispersion in Online Markets, in: Jansen, D. (2006): *The New Economy and Beyond: Past, Present and Future*, Edward Elgar Publishing, Northampton.

Baye, M. R., Morgan, J., & Scholten, P. (2004). Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site, *The Journal of Industrial Economics*, *52*(4), 463-496.

Beck, O., & Ülker, S.-L. (2022). Empirische Erforschung des Preismechanismus, *Wolfsburg Invisible Hand Studies*, WIHSt No. 8.

Borenstein, S., & Rose, N. L. (1994). Competition and Price Dispersion in the U.S. Airline Industry, *Journal of Political Economy*, *102*(4), 653-683.

Bridel, P. (1997). Money and General Equilibrium Theory: From Walras to Pareto (1870-1923), Edward Elgar Publishing, Cheltenham.

Brown, J. R., & Goolsbee, A. (2002). Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry, *Journal of Political Economy*, *110*(3), 481-507.

Burdett, K., & Judd, K. L. (1983). Equilibrium Price Dispersion, *Econometrica, 51*(4), 955-969.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers, *Management Science*, *46*(4), 563-585.

Chmielewski, L., & Kunzmann, O. (2021). Einführung in die empirische Forschung am Beispiel des Preismechanismus und der Bildung von Gleichgewichtspreisen, *Wolfsburg Invisible Hand Studies*, WIHSt No. 4.

Clar, F., Petrunina, J., Qitaku, A., & Zubke, L. (2021). Empirische Erforschung des Preismechanismus, *Wolfsburg Invisible Hand Studies*, WIHSt No. 3.

Dahlby, B., & West, D. S. (1986). Price Dispersion in an Automobile Insurance Market, *Journal of Political Economy*, *94*(2), 418-438.

Flemming, J., Boztepe, C., & Tawbe, S. (2021). Zur empirischen Preisforschung – Neoklassisches Marktmodell, *Wolfsburg Invisible Hand Studies*, WIHSt No. 6.

Hong, H., & Shum, M. (2006). Using price distributions to estimate search costs, *The RAND Journal of Economics*, *37*(2), 257-275.

Kessner, E., & Polborn, M. K. (2000). A new test of price dispersion, *German Economic Review*, *1*(2), 187-220.

Kornhardt, C. (2021). Empirische Untersuchung des neoklassischen Marktmodells anhand des Gleichgewichtspreises, *Wolfsburg Invisible Hand Studies*, WIHSt No. 1.

Lach, S. (2002). Existence and Persistence of Price Dispersion: An Empirical Analysis, *Review of Economics and Statistics*, *84*(3), 433-444.

Lewis, M. (2008). Price Dispersion and Competition with Differentiated Sellers, *Journal of Industrial Economics, 56*(3), 654-678.

Mankiw, N. G., & Taylor, M. P. (2020). Economics, 5th Edition, Cengage Learning EMEA, Andover.

Maynes, E. S. (1976). Decision-Making for Consumers: An Introduction to Consumer Economics, Prentice Hall, New York City/London.

Mises, L. von (1929). Kritik der Interventionismus, Untersuchungen zur Wirtschaftspolitik und Wirtschaftsideologie der Gegenwart, Fischer, Jena.

Mises, L. von (1940). Nationalökonomie, Theorie des Handelns und Wirtschaftens, Editions Union, Genf.

Möbius, D., Schmidt, M., & Waldhelm, S. (2021). Empirische Studie zum Thema Invisible Hand, *Wolfsburg Invisible Hand Studies*, WIHSt No. 5.

Moen, E. R., Wulfsberg, F., & Aas, Ø. (2020). Price Dispersion and the Role of Stores, *The Scandinavian Journal of Economic*, *122*(3), 1181-1206.

OECD (2021). Health at a Glance 2021: OECD Indicators, OECD Publishing, Paris.

Ötsch, W. O. (2019). Mythos Markt, Mythos Neoklassik: Das Elend des Marktfundamentalismus, Metropolis, Marburg.

Pindyck, R. S., & Rubinfeld, D. L. (2017). Microeconomics, 9th Edition, Pearson Education, London.

Pratt, J. W., Wise, D. A., & Zeckhauser, R. (1979). Price Differences in Almost Competitive Markets, *The Quarterly Journal of Economics*, *93*(2), 189-211.

Samuelson, P. A. (1948). Economics: An Introductory Analysis, McGraw-Hill, New York City.

Samuelson, P. A., & Nordhaus, W. D. (2009). Economics, 19th Edition, McGraw-Hill, New York City.

Scholten, P., & Smith, S. A. (2002). Price dispersion then and now: Evidence from retail and e-tail markets, *The Economics of the Internet and E-commerce, 11*, 63-88.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations, Strahan & Cadell, London.

Sorensen, A. T. (2000). Equilibrium Price Dispersion in Retail Markets for Prescription Drugs, *Journal of Political Economy*, *108*(4), 833-850.

Stigler, G. J. (1961). The Economics of Information, *Journal of Political Economy*, *69*(3), 213-225.

Van Hoomissen, T. (1988). Price Dispersion and Inflation: Evidence from Israel, *Journal of Political Economy*, *96*(6), 1303-1314.

Varian, H. R. (2014). Intermediate Microeconomics: A Modern Approach, 8th Edition, Norton, New York City.

Vukina, T., & Zheng, Z. (2010). Bargaining, Search, and Price Dispersion: Evidence from the Live Hogs Market, *Agricultural and Resource Economics Review*, *39*(3), 534-546.

Wahlers, J., & Schulenburg, S. von der (2022). Empirische Erforschung des Preismechanismus, *Wolfsburg Invisible Hand Studies*, WIHSt No. 9.

Walker, D. A. (1993). Walras's Models of the Barter of Stocks of Commodities. *European Economic Review*, *37*(7), 1425-1446.

Wenzlaff, L., & Leohold, S. (2022). Empirische Preisforschung – Neoklassisches Marktmodell, *Wolfsburg Invisible Hand Studies*, WIHSt No. 7.

Wildenbeest, M. R. (2011). An empirical model of search with vertically differentiated products, *The RAND Journal of Economics*, *42*(4), 729-757.

Yavuz, D. (2021). Der Preisbildungsprozess – Eine kritische Betrachtung der neoklassischen Theorie, *Wolfsburg Invisible Hand Studies*, WIHSt No. 2.

Younis, R., & Sokolowski, P. (2022). Empirische Studie zur Erforschung des Preismechanismus, *Wolfsburg Invisible Hand Studies*, WIHSt No. 10.

Ziegner, K., & Mützel, P. (2022). Empirische Erforschung des Preismechanismus, *Wolfsburg Invisible Hand Studies*, WIHSt No. 11.

Appendix A: Overview of the analyzed products and prices

Table A-1: Summary of price comparisons

Product	Price comparisons	Prices	WIHSt No.
Category "Drugstore"	**47**	**634**	**-**
Algemarina Trockenshampoo 200ml	2	9	6
Aptamil "Pronatura PRE" 800g	1	5	6
Bebe "Creme Intensivpflege" 50ml	3	14	6
Duschdas Sport 2-in-1 250ml	2	35	3, 5
Kneipp "Lebensfreude"	3	32	4
Head&Shoulders Apple Fresh 300ml	1	13	10
Head&Shoulders Classic Clean 300ml	2	40	8
Hipp "Ultra Sensitiv" Feuchttücher 4er Pack	1	4	6
Hipp "Zart Pflegend" Feuchttücher 4er Pack	1	4	6
I Love Extreme Mascara "Volume"	3	15	6
Nivea Deoroller Fresh pure 0%	3	25	11
Nivea Dry Impact Deo 150ml	1	7	5
Nivea Soft 200ml	3	29	10
Odol-med3 Zahnpasta Extra White 125ml	1	20	10
Pampers "Baby Dry" 21 Stück	1	3	6
Pampers "Premium Protection" 26 Stück	3	24	6
Pantene PRO-V Repair & Care 300ml	1	10	7
Pflaster Hansaplast "Classic"	3	44	4
Pril Kraftgel Ultra Plus	1	19	9
Schauma 7 Kräuter Shampoo	1	17	9
Tempo Taschentücher 30 x 10 Stück	1	22	10
UHU Kleber 21g	6	201	2
Zahnpasta Elmex "Kariesschutz"	3	42	4
Category "Food"	**59**	**776**	**-**
Airwaves Kaugummis Cool Cassis 12 Stück	1	22	10
Airwaves Strong Kaugummi 12 Stück	1	9	3
Barilla Penne Rigate 500g	1	6	5
Coca-Cola Original Taste 0,33l	1	17	10
Dr. Oetker Ristorante Salame 320g	2	25	5, 9
Professional White Kaugummi 50 Stück	1	12	3
Funny-frisch "Ungarisch" 175g	2	24	1, 3
Géramont "Classic" 200g	1	6	1
Haribo Goldbären 200g	2	40	8
Haribo Happy Cola 200g	1	10	3
Heineken Pils 6 x 0,33l	1	10	1
Honig Langnese "Flotte Biene" 250g	3	19	4
Jägermeister 0,7l	1	9	5
Konfitüre Schwartau Extra Erdbeere 340g	3	21	4
Leibniz Keks'N Crem Choco 228g	1	9	3
Lindt Lindor Kugel Milch 100g	1	11	3
Maggi Ravioli in Tomatensauce 800g	1	15	9
Maggi Würze 250g	1	20	8
Milka "Haselnussschokolade" 100g	1	6	1
Milka Alpenmilch 100g	3	50	5, 8

Milka Luflée Schokolade 100g	1	9	7
Niemand Dry Gin 0,5l	6	101	2
Nutella Nuss-Nugat-Creme 450g	4	32	3, 4
Pom-Bär Original 75g	1	32	3
Pringles "Original" 200g	2	21	1, 5
Pringles Chips Sour Cream & Onion 200g	1	13	3
Red Bull Classic 250ml	5	53	1, 3, 11
Red Bull Sugarfree 250ml	1	19	9
Ritter Sport Alpenmilch 100g	2	40	8
Ritter Sport Voll-Nuss 100g	1	28	3
Snickers 50g	2	16	2
Toffifee 125g	1	29	3
Uncle Ben's Express Langkornreis 250g	2	36	8
Vilsa Classic 12 x 0,7l Kasten	1	6	5
Category "Other"	**40**	**807**	**-**
AirPods 2. Gen. / MV7N2ZM/A	6	338	2
Aspirin 500mg (20 Tabletten)	2	18	5
Baby Einstein Magic Touch Piano	2	64	7
Big Bobby Car Classic Sansibar	2	18	7
Converse Chuck Taylor All Star High	3	29	11
FIFA 21 (Playstation 4)	1	39	1
FIFA 22 (Playstation 5)	1	11	5
HP 302 Cyan/Magenta/Gelb Druckerpatrone	3	35	11
JBL Flip 5	1	30	1
KTM Radical Kids Training Bike	2	13	7
Nintendo Switch	1	37	1
PS4 Wireless Dualshock Controller, V2	1	43	1
Sony Playstation 5 Disc Version	1	5	6
TomTom "Go Discover 7"	2	14	6
Toniebox Starterset inkl. Kreativtonie	1	11	7
UNO Standard	3	31	11
WMF Kult X Mix & Go 0,6l	5	47	5, 11
WMF Toaster Stelio Edelstahl	1	15	9
Xiaomi Scooter 1S	2	9	6
Total	**146**	**2,217**	**-**

Appendix B: Translation of product descriptions

Table A-2: Product descriptions in English

Product name	English description
Category "Drugstore"	
Algemarina Trockenshampoo 200ml	Dry shampoo
Aptamil "Pronatura PRE" 800g	Baby food
Bebe "Creme Intensivpflege" 50ml	Moisturizer
Duschdas Sport 2-in-1 250ml	Shower gel
Kneipp "Lebensfreude"	Shower gel
Head&Shoulders Apple Fresh 300ml	Shampoo
Head&Shoulders Classic Clean 300ml	Shampoo
Hipp "Ultra Sensitiv" Feuchttücher 4er Pack	Wet wipes
Hipp "Zart Pflegend" Feuchttücher 4er Pack	Wet wipes
I Love Extreme Mascara "Volume"	Mascara
Nivea Deoroller Fresh pure 0%	Deodorant stick
Nivea Dry Impact Deo 150ml	Deodorant spray
Nivea Soft 200ml	Moisturizer
Odol-med3 Zahnpasta Extra White 125ml	Toothpaste
Pampers "Baby Dry" 21 Stück	Diapers
Pampers "Premium Protection" 26 Stück	Diapers
Pantene PRO-V Repair & Care 300ml	Hair care product
Pflaster Hansaplast "Classic"	Plaster
Pril Kraftgel Ultra Plus	Dishwashing detergent
Schauma 7 Kräuter Shampoo	Shampoo
Tempo Taschentücher 30 x 10 Stück	Tissues
UHU Kleber 21g	Glue
Zahnpasta Elmex "Kariesschutz"	Toothpaste
Category "Food"	
Airwaves Kaugummis Cool Cassis 12 Stück	Chewing gum
Airwaves Strong Kaugummi 12 Stück	Chewing gum
Barilla Penne Rigate 500g	Pasta
Coca-Cola Original Taste 0,33l	Soft drink
Dr. Oetker Ristorante Salame 320g	Pizza
Professional White Kaugummi 50 Stück	Chewing gum
Funny-frisch "Ungarisch" 175g	Potato chips
Géramont "Classic" 200g	Cheese
Haribo Goldbären 200g	Jelly sweets
Haribo Happy Cola 200g	Jelly sweets
Heineken Pils 6 x 0,33l	Beer
Honig Langnese "Flotte Biene" 250g	Honey
Jägermeister 0,7l	Liquor
Konfitüre Schwartau Extra Erdbeere 340g	Jam
Leibniz Keks'N Crem Choco 228g	Cookies
Lindt Lindor Kugel Milch 100g	Chocolate
Maggi Ravioli in Tomatensauce 800g	Pasta
Maggi Würze 250g	Sauce
Milka "Haselnussschokolade" 100g	Chocolate
Milka Alpenmilch 100g	Chocolate
Milka Luflée Schokolade 100g	Chocolate

Niemand Dry Gin 0,5l	Liquor
Nutella Nuss-Nugat-Creme 450g	Hazelnut spread
Pom-Bär Original 75g	Potato chips
Pringles "Original" 200g	Potato chips
Pringles Chips Sour Cream & Onion 200g	Potato chips
Red Bull Classic 250ml	Energy drink
Red Bull Sugarfree 250ml	Energy drink
Ritter Sport Alpenmilch 100g	Chocolate
Ritter Sport Voll-Nuss 100g	Chocolate
Snickers 50g	Chocolate bar
Toffifee 125g	Caramel candy
Uncle Ben's Express Langkornreis 250g	Rice
Vilsa Classic 12 x 0,7l Kasten	Mineral water
Category "Other"	
AirPods 2. Gen. / MV7N2ZM/A	Earphones
Aspirin 500mg (20 Tabletten)	Medicine
Baby Einstein Magic Touch Piano	Piano
Big Bobby Car Classic Sansibar	Toy car
Converse Chuck Taylor All Star High	Shoes
FIFA 21 (Playstation 4)	Video game
FIFA 22 (Playstation 5)	Video game
HP 302 Cyan/Magenta/Gelb Druckerpatrone	Printer cartridge
JBL Flip 5	Portable speaker
KTM Radical Kids Training Bike	Bike
Nintendo Switch	Game console
PS4 Wireless Dualshock Controller, V2	Game controller
Sony Playstation 5 Disc Version	Game console
TomTom "Go Discover 7"	Navigation device
Toniebox Starterset inkl. Kreativtonie	Toy
UNO Standard	Card game
WMF Kult X Mix & Go 0,6l	Blender
WMF Toaster Stelio Edelstahl	Toaster
Xiaomi Scooter 1S	Electric scooter

Appendix

Additional Contributions to the Scientific Debate

During the course of my doctoral studies, I have made further contributions to the scientific community, which are listed here.

In a multidisciplinary research team with experts from the research fields of law, computer sciences, and innovation, I have contributed to a research paper on the application of artificial intelligence in the mobility sector and emerging legal and ethical issues. The paper was published through *Verlag C.H. Beck.* In particular, I helped conceptualize the research, organized the literature, wrote the chapter on the organizational dimension, and took care of formatting.

- Bittner, J., Debowski, N., Lorenz, M., Raber, H. G., Steege, H., & Teille, K. (2021). Recht und Ethik bei der Entwicklung von Künstlicher Intelligenz für die Mobilität, *Neue Zeitschrift für Verkehrsrecht, 34(*10), 505-514.

Moreover, the following presentations and workshops were held based on the results of this dissertation:

- Future Learning Session: Algorithm Aversion - Warum wir Zweifel haben (presentation and discussion with over 300 Volkswagen employees together with research partner Jan René Judek), April 2021.
- The Tragedy of Algorithm Aversion (presentation and discussion), 2021 ESA Global Online Around-the-Clock Conference, July 2021.
- Gestaltungswille und Algorithm Aversion – Die Auswirkungen der Einflussnahme im Prozess der algorithmischen Entscheidungsfindung auf die Algorithm Aversion (presentation and discussion), Jahrestagung der Gesellschaft für experimentelle Wirtschaftsforschung, Magdeburg, October 2021.
- An Introduction to Algorithm Aversion (presentation and discussion with 145 Volkswagen employees at a company-wide online conference), January 2022.
- Method Tryout: Experimentelle Forschung (interactive hands-on workshop on experimental economics with Volkswagen employees), February 2022.
- Impact of the Decoy Effect on Algorithm Aversion (presentation and discussion with Volkswagen managers), June 2022.

Declaration of Contribution to Each Study

First contribution: Reducing Algorithm Aversion through Experience
45% - Conceptualization, literature analysis, methodology, experimental design, programming and conducting of the economic experiment, data analysis, writing, formatting, editing.

Second contribution: The Tragedy of Algorithm Aversion
45% - Conceptualization, literature analysis, methodology, experimental design, programming and conducting of the economic experiment, data analysis, writing, formatting, editing.

Third contribution: Comparing Different Kinds of Influence on an Algorithm in Its Forecasting Process and Their Impact on Algorithm Aversion
45% - Conceptualization, literature analysis, methodology, experimental design, programming and conducting of the economic experiment, data analysis, writing, formatting, editing.

Fourth contribution: Impact of the Decoy Effect on Algorithm Aversion
100% - Conceptualization, literature analysis, methodology, experimental design, programming and conducting of the economic experiment, data analysis, writing, formatting, editing.

Fifth contribution: Algorithm Aversion as an Obstacle in the Establishment of Robo Advisors
45% - Literature analysis, methodology, experimental design, programming and conducting of the economic experiment, data analysis, writing, formatting, editing.

Sixth contribution: Interest Rate Forecasts in Latin America
45% - Conceptualization, literature analysis, methodology, research design, data collection, data preparation, data analysis, writing, formatting, editing.

Seventh contribution: Sticky Stock Market Analysts
45% - Conceptualization, literature analysis, methodology, research design, data collection, data preparation, data analysis, writing, formatting, editing.

Eighth contribution: Unicorn, Yeti, Nessie, and Neoclassical Market – Legends and Empirical Evidence
45% - Conceptualization, literature analysis, methodology, research design, data collection, data preparation, data analysis, writing, formatting, editing.

September 1st, 2022, Marco Lorenz

Assurance According to §12 PStO

Ph.D. program in Economics

Declaration for admission to the doctoral examination

I confirm

1. that the dissertation that I submitted

Algorithm Aversion and Other Causes of Bias in Decision Behavior *- Studies on Algorithm Aversion, Capital Market Forecasting, and Price Dispersion*

was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,

2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,

3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,

- No remuneration or equivalent compensation were provided
- No services were engaged that may contradict the purpose of producing a doctoral thesis

4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

_____

September 1st, 2022, Marco Lorenz