FINANCIAL MARKET ACTORS:
COGNITIVE BIASES, PORTFOLIO DIVERSIFICATION
AND FORECASTING ABILITY

Dissertation

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

submitted by

Thomas Nahmer

born in Wolfsburg

Göttingen, 2019

CONTENTS

CHAPTER 5

(with Ibrahim Filiz and Markus Spiwoks)

Submitted to:

Journal of Behavioral and Experimental Finance

Published in:

Wolfsburg Working Papers, 19-02, February 2019.

CHAPTER 6

(with Ibrahim Filiz, Markus Spiwoks and Kilian Bizer)

Submitted to:

Applied Economics

Published in:

Wolfsburg Working Papers, 18-02, June 2018.

Diskussionsbeiträge der Sonderforschungsgruppe Institutionenanalyse, 18-2, August 2018.

Chapter 1

INTRODUCTION AND SUMMARY

"The story of the sciences in the twentieth century is one of a steady loss of certainty. Much of what was real and machine-like and objective and determinate at the start of the century, by mid-century was a phantom, unpredictable, subjective and indeterminate. What had defined science at the start of the century—its power to predict, its clear subject/object distinction—no longer defined it at the end." (Arthur, 1994)

This certainly - and particularly - also applies to economic sciences. In the 19th century, the economic debate was still characterized by a remarkable degree of diversity. Alongside marginalism there was the historical school of economics, socialist economics or the romantic school of economics. By the beginning of the twentieth century, Menger, Böhm-Bawerk, Wicksell and Walras had prevailed against Schmoller and Wagner, against Marx and Engels, and against Adam Müller. The mathematical, model-based penetration of economic life, so-called neo-classicism, assumed that it was fitting to describe individuals as rational utility maximisers (homo oeconomicus) and that the interactions between these individuals could be accurately reflected by the market model. However, the faith of neo-classicist economists in the market was severely shaken in 1929 by Black Friday and the subsequent Great Depression. A few years later, Keynes (1936) showed that it is by no means true that markets always act rationally under all circumstances. Nevertheless, this was not the end of neoclassical thinking, which later experienced a revival from the end of the 1960s onwards.

However, there were always critics of neoclassical theory, who partly developed some very interesting alternative approaches. New institutional economics is certainly one of these, particularly the aspects of transaction cost theory and the theory of asymmetric information. From the mid-1970s, they were joined by behavioral economics as a new branch of economic research. This is closely related to two names: Amos Tversky and Daniel Kahneman, and began with their publications "Judgment under Uncertainty: Heuristics and Biases" (1974) and "Prospect Theory: A Study of Decision Making Under Risk" (1979). Prospect theory was the first comprehensive and theoretically sound outline of a descriptive decision theory, and finally ended the hegemony of normative rationality. The systematic description of framing effects and changing risk behavior dependent on the translation of a decision into profits and losses yielded important new recognitions (Thaler, 1980). Based on the findings of Kahneman and Tversky, the academic discourse over the past 25 years has increasingly focused on explanatory models founded upon behavioral science. In empirical studies, a large number of obvious anomalies and inconsistencies were revealed which challenged the market efficiency paradigm, and in many cases clearly refuted it.

The increasing frequency of severe disruptions in the capital markets (1987, 2000 and 2008) has further fueled existing doubts about neoclassical theory and increased the interest in alternative research approaches. Today, economic research is characterized by a pleasing variety of both theories and methodology, and empirical research has emancipated itself from model theories of economic behavior. Experimental research has become an integral element of economics: alongside psychological aspects, cultural, sociological and political factors are also taken into account. In the field of neuroeconomics even medical procedures are involved in economic research.

This cumulative dissertation also reflects the new methodological diversity in economics. Three out of the five studies here consider the decision-making behavior of real subjects with the aid of laboratory experiments. The first study analyses real investment alternatives with the aid of simulated calculations. In the final study, large amounts of real data (capital market forecasts by experts) are evaluated econometrically.

In terms of content, the dissertation deals with capital markets. In spite of a wide range of attempted interpretations, sophistic market theories and ambitious models, academic research still has difficulty in explaining the fundamental characteristics of real capital markets. Capital markets are complex systems with a considerable degree of momentum, and they are driven by a large number of heterogeneous market participants who possess incomplete information and a limited capacity for rational decision-making (Rapp und Cortes, 2017).

For investments in the capital markets, portfolio theory (Markowitz, 1952) plays a significant role and continues to form an important basis for decisions on the structuring of investment portfolios. For risk-averse investors, diversifying the contents of their portfolio is a meaningful strategy. This applies whenever the direction of future price trends is uncertain, because diversification is pointless when market movements are highly predictable. In that case the optimal strategy would be to invest solely in the security with the highest expected increase in value (Markowitz, 1991). In practice, however, investors continue to hold underdiversified portfolios which contradict the basic tenet of portfolio theory.

The first three studies of this cumulative dissertation address different aspects of portfolio management. In the first study, a simulated calculation is used to examine the suitability of fine wine as a means of diversification. In the second study, possible causes for sub-optimal diversification decisions are analyzed. And in the third study, a method to improve the measuring of the risk preference of subjects is presented. The fourth study deals with the influence of mood on the tendency towards herd behavior in the context of share price forecasts. Finally, the fifth study evaluates real interest rate forecasts for the Asia-Pacific region in order to be able to assess the forecasting skills of the financial analysts in question. In the following sections I present a more detailed look at the five studies:

First contribution – Investing in Fine Wine from the Perspectives of Diversification and Costs

For centuries now, investors have been practicing the concept of spreading their financial investments widely. By investing in securities of different geographical origins, from various issuers and in different currencies, the risk level of a portfolio is reduced. The theoretical basis of portfolio theory was laid down by Harry Markowitz (Markowitz, 1952). By diversifying the assets held, the risk of a portfolio can be reduced while maintaining the same expected return, or returns can be increased without also increasing the level of risk involved. This is dependent on acquiring a combination of investments which are not completely correlated with each other.

In the aftermath of the global financial crisis in the years 2008-2009, interest rates fell significantly worldwide. This is one of the main reasons why investments in alternative asset classes have recorded such impressive growth rates. Alongside hedge funds, private equity and commodities, so-called exotics or collectables are also increasingly being considered as investment opportunities. Fine wine is also a commodity whose positive diversification features have been noted. The foundation for this is that the factors which influence wine prices differ significantly from those which influence the prices of classical asset classes. In this way, an important precondition for meaningful diversification is already fulfilled.

Previous studies which have taken a closer look at the suitability of fine wine for diversification purposes based their arguments on index data. However, making calculations on the basis of index data contains pitfalls: index data does not include any costs. And precisely this question is of particular relevance for investments in fine wine, as there have been no low-cost index funds until now.

In order to create a direct link to previous studies, the first stage of the analysis was carried out on an index basis. And then, in a second stage based on real-life investment opportunities, the actual investment costs were taken into account. The reference currencies were the US dollar and the euro. For the analysis from an equity index perspective, the MSCI World Index was used, and for bonds the JPM Global Government Bond Index was deployed. Regarding the data for investment in fine wine, the main focus was on the Liv-ex-50 Index calculated by the London wine exchange Liv-ex, which was founded in 1999. The period of observation was January 2004 to May 2018. In the case of the real investments, index funds were used for the data analysis of equities and bonds. As there is no index fund for fine wine, the Liv-ex-50 index was used, including all of the costs of a real investment. The cost elements for the calculation were made available by the Liv-ex for the period from March 2010 onwards. The period of observation was thus March 2010 to May 2018.

Various portfolio compositions were compared over the periods indicated. On the one hand, a portfolio of 50% shares and 50% bonds was compared to a portfolio of 45% shares, 45% bonds and 10% fine wine. On the other hand, a portfolio of 25% shares and 75% bonds was compared to a portfolio of 20% shares, 70% bonds and 10% fine wine. As benchmarks, the annualized return, the standard deviation and the Sharpe ratio of the respective portfolios were calculated.

The findings are sobering and clearly contradict those of the previous studies. At an index level, the inclusion of fine wine merely leads to a slight improvement of the annualized return, but at the same time to an increase in risk. Only in the case of one portfolio in euros was the Sharpe ratio somewhat better after the inclusion of 10% fine wine. When considering the real investment opportunity, the considerable costs of an investment in fine wine have to be viewed against the low costs of an index fund for equities and bonds. In US dollars and euros, the annualized returns were lower in comparison to the portfolios without the inclusion of fine wine, and at the same time the risk levels are higher. The inclusion of wine always led to a decrease in the Sharpe ratio.

The conclusion on the diversification properties of fine wine in a portfolio which is already broadly diversified in traditional asset classes can be fittingly described as sobering. Fine wine does not lead to any improvement of the Sharpe ratio from the index perspective in US dollars, and when the base currency is euros only a slight increase was achieved in one out of two portfolios. When all costs are taken into account, there was not a single case where the risk-return profile of a portfolio was improved by the inclusion of fine wine. However, on the positive side it can be noted that the costs of an actual investment in fine wine have been noticeably reduced. This is due to the establishment of the Liv-ex global fine wine market, which charges considerably lower transaction fees than classical auction houses.

Second contribution – Portfolio Diversification: The Influence of Herding, Status-Quo Bias and Gambler's Fallacy

Empirical and experimental research has provided a multitude of explanations for why subjects obviously find it so difficult to make meaningful decisions when it comes to diversifying their portfolios. The explanations range from difficulties in dealing with correlations to wrongly-assessed probabilities of occurrence, the so-called illusion of knowledge or a simple home bias.

Other possible causes for sub-optimal diversification decisions were established in interviews carried out with experts. Observing the investment decisions of others seems to make some investors lose sight of optimal diversification opportunities (herding behavior). The statements and recommendations of well-known investors can play a significant role here (guru effect). In addition, clinging to existing portfolio compositions can also prevent an optimal level of diversification (status-quo bias). Finally, an orientation towards what are presumed to be patterns in share prices (gambler's fallacy) can lead to wayward investment decisions.

In this context, the question arises as to whether herding behavior, status-quo bias and/or gambler's fallacy are really suitable explanations for why many subjects fail to diversify their portfolios sufficiently.

The significance of these phenomena in economic decision-making situations has been empirically proven, but in the context of diversification decisions, they have not been analyzed by means of experiments until now.

In this experiment, the subjects only have two investment alternatives to choose from (share A and share B). A total of four units can be invested. These two shares can only exhibit two price movements, and both events have a probability of occurrence of 50%. The return on share A and the return on share B are independent random events. With regard to their expected returns and risk exposure, share A and share B do not differ. In this situation the efficient frontier is reduced to a single point (a 50-50 mix of shares A and B) so that the exact nature of the risk aversion of the subjects exerts no influence on the optimal decision with regard to the portfolio.

In Treatment 1, in each round the subjects receive all of the information available about their fellow players: they receive a performance table of the participants along with the decisions they have made about their portfolios. In Treatment 2 the subjects are not informed about the investment decisions of the other subjects and their investment success. The comparison between Treatment 1 and Treatment 2 is intended to provide information about how strongly herding behavior (orientation towards the majority of their fellow players or towards the most successful player (guru)) keeps the subjects from optimally diversifying their portfolios.

In Treatment 1, the subjects begin with different sub-optimal investment portfolios. In Treatment 3, all of the subjects receive the optimal portfolio as an initial starting point. If the phenomenon of status-quo bias occurs, the optimal portfolio should be chosen more frequently in Treatment 3, and the average risk exposure should be lower than in the basic treatment.

Given that the rational strategy of choosing two units each of share A and share B only applies to risk-averse investors, non-risk averse subjects are identified using the method of Holt and Laury (2002) and not taken into consideration in the remainder of the evaluation.

Overall, it had to be noted that in this experiment to the large majority of portfolio decisions (more than 62%) have to be considered sub-optimal. This is also shown by the fact that the average standard deviation of the chosen portfolios is higher than that of the standard deviation of the portfolio which would have been chosen if the rational strategy had been pursued.

The portfolio decisions of the subjects remained fragmented throughout the entire game. A tendency towards herding was not recognizable. Neither the most frequently chosen portfolios from the previous round nor the portfolio of the leading player had a decisive effect on the subjects' decisions.

Status-quo bias did not exert an important influence on the occurrence of sub-optimal diversification decisions either. The differing starter portfolios in Treatments 1 and 3 did not have a lasting effect on the composition of the subjects' portfolios.

However, the observation of randomly-occurring presumed patterns in the returns of the two shares did indeed have a very significant influence on the portfolio decisions of the subjects. This was also confirmed by the statements of the subjects who were interviewed in rounds 5 and 11 about the motives behind their decisions: over 40% stated that the identification of patterns of results had formed the basis for their decisions. It can thus be stated that gambler's fallacy can play an important role in sub-optimal portfolio diversification.

Third contribution – Measurement of Risk Preference

Individual risk preference is a key element in economic theory on decision-making under uncertainty. Experimental research on diversification behavior also requires a clear differentiation between risk-averse, risk-neutral and risk-loving subjects, because decisions which can be absolutely meaningful for a risk-loving subject are often completely inconceivable for a risk-averse subject and vice versa. Robust findings in experimental research on diversification can only be obtained when one knows how to categorize the risk preferences of the subject.

Well-known procedures for the measurement of risk preference are put to the test here, and a new approach is presented which is superior to the previous ones with regard to the unequivocal identification of risk preference while taking a relevant risk of loss into account.

A good procedure for determining risk preference should above all comply with three criteria: it must be a simple and clear procedure; it must be possible to unambiguously differentiate between risk-averse, risk-neutral and risk-loving subjects; and the influence of loss aversion on risk preference must not be neglected. It should also be easy to understand so as not to confuse the subjects. They should be able to grasp and assess the decision-making situation without difficulty. Clearly identifying risk preference is necessary in order to correctly interpret the results of experimental research on diversification behavior. And finally, the possibility of not inconsiderable losses is important, as this is an essential characteristic of real capital markets, and loss aversion plays a significant role in the formation of risk preferences.

Among the existing approaches, that used by Holt and Laury (2002) has undoubtedly received the most attention. The fact that not every subject - given the complexity of the task - makes decisions which lead to unambiguous results has already been established several times (Jacobson und Petrie, 2009; Charness und Viceisza, 2011). The approach used by Eckel and Grossman (2008) is significantly simpler and clearer, and that deployed by Crosetto and Filippin (2013) even more so. However, all three procedures exhibit the weakness that in certain situations it is not possible to differentiate in an unambiguous and reliable way between risk-averse, risk-neutral and risk-loving subjects. In addition, in these three approaches the influence of loss aversion on risk preference is not taken into consideration, or not sufficiently.

The procedure proposed here, however, fulfils all three of the criteria mentioned above. It deals with a decision to choose between two lotteries. The subjects take a card - they can choose between taking a card from pile A or one from pile B. Both piles consist of four playing cards each. The subjects are informed that the expected return is identical in both piles. In addition, the subjects are made aware of the fact that pile A leads to results which fluctuate slightly around the expected value (risk-averse subjects choose this option), while pile B leads to results which fluctuate considerably around the expected value (risk-loving participants choose this option). In addition, there is the option for subjects to be indifferent towards choosing pile A or pile B (risk-neutral participants choose this option).

The game is played using three treatments which have the same expected value and only differ with regard to their level of risk. Treatment 1 does not contain any possibility of losses, while in Treatment 2 there is a possibility of a small loss and in Treatment 3 there is a possibility of a significant loss. The reason behind the selection of the treatments is the expectation that only the possibility to suffer a relevant loss will lead to a correct identification of the risk preferences of the subjects.

The results were clear and largely in line with our expectations. Comparing Treatment 1 (no possibility of loss) and Treatment 2 (possibility of a small loss), there was no significant difference between the number of participants who chose pile A or pile B. In treatment 3, however (possibility of a significant loss) a marked influence on risk preferences can be noted. In Treatment 3 less than half of the subjects (in percentage terms) chose the risky variation (pile B)

than was the case in Treatments 1 and 2. Our presumption that a high risk of loss influences the risk preference of subjects considerably more than a low risk of loss was thus confirmed.

Overall, it can be stated that taking a possibility of a significant loss into account leads to a more realistic recording of the three categories of risk preference (risk averse, risk neutral und risk loving). In the form of our treatment 3 we are thus proposing a new approach which discriminates clearly between risk-averse, risk-neutral and risk-loving subjects. In addition, the approach is clear and simple and contains a possibility to suffer a significant loss.

Fourth contribution – Herd Behavior and Mood

In economics, herding behavior is considered to be present when subjects imitate the behavior of other subjects, or make economic decisions on the basis of the views, assessments or actions of others. By now it is considered proven that the mood of subjects has an influence on their economic decision-making. There have been a large number of studies on this topic. However, until now no-one has conducted experiments on whether different moods also have an affect on the likelihood of herding behavior.

In our experiment, the subjects forecast share prices. They have the opportunity to determine future share price levels via four fundamental influencing factors which have a constant effect on the share price. An error-free forecast is made more difficult by a random influence which leads to a situation where even subjects who act rationally only manage to successfully forecast an average of 40% of share prices. The greater the distance between the forecasts of the subjects and the expected value of the share (rational strategy), the lower the probability that their forecast will prove correct. In extreme cases, the probability that the forecast is correct can fall to 0%.

After they have made their preliminary forecasts, the subjects are informed of the consensus forecast and thus offered the chance to orientate themselves towards other subjects and to possibly exhibit herding behavior. In a between-subjects design, three treatments are used (neutral, positive and negative mood). Mood is influenced by means of film excerpts which are shown to the subjects.

As expected, the subjects did not act in accordance with homo oeconomicus, a figurative human characterized by the ability to make rational decisions: they did not always choose the rational strategy. Less than a third of all forecasting decisions followed the rational strategy, and less than a tenth of all subjects made rational decisions throughout all rounds of the game. After the consensus forecast had been announced, the subjects adjusted their forecasts in more than a third of all cases. These adjustments were almost exclusively in the direction of the consensus forecast. A clear tendency towards herding behavior was thus observed. The average standard deviation of the preliminary forecasts was significantly higher than the average standard deviation of the final forecasts. This is also a sign of herding behavior. We observed hardly any indications of anti-herding, however: only one in 29 of the revisions of forecasts were not carried out in the direction of the consensus forecast, but in the opposite direction. The rational strategy was followed significantly more often in the treatment positive

and treatment negative than in the treatment neutral. Adjustments of the forecasts in the direction of the consensus forecast (herding behavior) occurred significantly less often in the treatments positive and negative than in the treatment neutral.

Herding behavior can be observed very frequently. The mood of the subjects has an influence on the occurrence of herding: the tendency towards herding behavior is stronger in a neutral mood than in a positive or negative mood. In addition, less attention is paid to the rational strategy in a neutral mood than in a positive or negative mood.

Fifth contribution – The Accuracy of Interest Rate Forecasts in the Asia-Pacific Region: Opportunities for Portfolio Management

Future interest rate trends are of key significance for almost all investment decisions on the capital markets. This applies not only to price trends on the bond markets, but also to share prices (discounting rates) and exchange rates (interest rate parity theory). As a rule, financial market analyses normally begin with a forecast of future interest rate trends.

It is therefore not surprising that the reliability of interest rate forecasts has been of great interest to academics and businesspeople for a considerable time now. In the past 40 years, a large number of empirical studies on the reliability of interest rate forecasts have been presented. The results varied, but were predominantly disappointing. Until today, forecasts of US interest rates have formed the main focus of studies, although European interest rates, particularly those in the UK and Germany, have also been analyzed. In the Asia-Pacific region, previous studies have largely focused on Japan. Otherwise there have been very few publications. At the same time, the economic significance of the Asia-Pacific area has increased considerably: the region accounts for 60% of world economic growth, and the weighting of the Asia-Pacific bond market in the global bond indices already exceeds 21%.

The study on interest rate forecasts in the Asia-Pacific region covers the period from 1990 to 2015 and focused on interest rate forecasts from Australia, China, Hong Kong, India, Indonesia, Malaysia, New Zealand, Singapore, South Korea, Taiwan and Thailand. As a basis we used forecasting data which had been published on a monthly basis in the journal Asia Pacific Consensus Forecasts. In addition to the analysis of consensus forecasts, all of the forecast time series issued by banks, investment companies, consulting firms, associations and industrial companies were evaluated. This amounted to a total of 532 forecast time series with 85,264 individual interest rate forecasts. The deployment of a large number of procedures to measure the quality of forecasts enabled us to create a comprehensive picture of forecasting performance in the Asia-Pacific region. Comparisons to the naïve forecast were carried out with the aid of the Diebold-Mariano test. In order to assess the quality of the forecasts with regard to the direction of the interest rate trends, the sign accuracy test was used, while the unbiasedness test enabled a search for systematic forecasting errors. Finally, the existence of topically-oriented trend adjustments was tested for with the aid of the TOTA coefficient.

The results were very sobering in part. A very large proportion of all forecast time series are characterized by the phenomenon of topically-oriented trend adjustments. This means that

the overwhelming majority of all forecast time series tend to reflect the present rather than the future. In addition to this, almost all of the forecast time series proved to be distorted.

However, some of the results of the study are also surprisingly positive. The sign accuracy test revealed that almost half of the forecast time series grasped the future trend (rising or falling interest rates) significantly better than a random walk forecast. In this context, at least part of the forecasts for Australia, China, Hong Kong, India, Malaysia, New Zealand, Singapore and Thailand proved to be particularly successful.

Compared to the naïve forecast, some of the interest rate forecasts analyzed, particularly those for interest rates in Hong Kong, India, Indonesia, Malaysia and New Zealand were surprisingly positive.

Overall it can be stated that - at least in some countries and for some forecast horizons - forecasts of future interest rate trends in the Asia-Pacific region are significantly more successful than those made in other parts of the world. This has consequences for portfolio management: it opens up opportunities to achieve excess returns by investing selectively.

References

Arthur, W. Brian (1994): The End of Certainty in Economics, in: Einstein Meets Magritte: An Interdisciplinary Reflection on Science, Nature, Art, Human Action and Society vol. 1, Emma, 255-265.

Charness, G. / Viceisza, A. (2011): Comprehension and Risk Elicitation in the field: Evidence From Rural Senegal, IFPRI Discussion Papers, 1135.

Crosetto, P. / Filippin, A. (2013): The "Bomb" Risk Elicitation Task, Journal of Risk Uncertainty, 47(1), 31-65.

Eckel, C. C. / Grossman, P. J. (2008): Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices, Journal of Economic Behavior and Organization, 68(1), 1-17.

Holt, C. A. / Laury, S. K. (2002): Risk Aversion and Incentive Effects, American Economic Review, 92(5), 1644-1655.

Jacobsen, S. / Petrie, R. (2009): Learning from Mistakes: What Do Inconsistent Choices over Risk Tell Us? Journal of Risk and Uncertainty, 38(2), 143-158.

Keynes, J. M. (1936): The General Theory of Employment, Interest and Money, Macmillan, London.

Markowitz, H. M. (1952): Portfolio Selection, Journal of Finance, 7(1), 77-91.

Markowitz, H. M. (1991): Foundations of Portfolio Theory, Journal of Finance, 46(2), 469-477.

Rapp, H.-W. / Cortes, A. (2017): Cognitive Finance: Neue Sicht auf Wirtschaft und Finanzmärkte, SpringerGabler.

Thaler, R. H. (1980): Toward a Positive Theory of Consumer Choice, Journal of Economic Behavior & Organization, 1(1), 39–60.

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

Tversky, A. / Kahneman, D. (1979): Prospect Theory: A Study of Decision Making Under Risk, Econometrics, 47, 263-291.

Chapter 2

INVESTING IN FINE WINE FROM THE PERSPECTIVES OF DIVERSIFICATION AND COSTS

Abstract

This paper examines the meaningfulness of fine wine as an alternative investment, with particular focus on the costs of investing in fine wine. Is fine wine suitable for further diversifying and thus improving the risk-return profile of portfolios invested in global equities and bonds? This analysis takes place in an initial stage on an index basis and in a second stage on the basis of real investment opportunities. The reference currencies are the US dollar and the euro. In order to observe stock indexes, the MSCI World Index is used, and for bonds the JPM World Government-Bond Index is deployed. Regarding the data for investment in fine wine, the main focus is on the Liv-ex-50 Index. The time period is defined by the availability of the data. For the observation of indices, the period is from the beginning of 2004 to May 2018. For observation on the basis of a real investment the period is from March 2010 to May 2018. In the case of the real investment, index funds are used for the data analysis of equities and bonds. As there is no index fund for fine wine, the Liv-ex-50 index is used including all of the costs of a real investment.

Various portfolio compositions are used for the periods indicated. On the one hand, a portfolio of 50% equities and 50% bonds is compared to a portfolio of 45% equities, 45% bonds and 10% fine wine. On the other hand, a portfolio of 25% equities and 75% bonds is compared to a portfolio of 20% equities, 70% bonds and 10% fine wine. As benchmarks, the annualised return, the standard deviation and the Sharpe ratio of the respective portfolios are calculated.

The results for the periods indicated are sobering. The inclusion of fine wine leads - at an index level - to only a slight improvement of the annualised return, but to a marked increase in risk. When considering the real investment, the considerable costs of an investment in fine wine come to bear. The annualised return is lower and at the same time the risk is higher than that of portfolios which do not include fine wine. It is only when the index is viewed in euros that a slight improvement of the Sharpe ratio in one portfolio can be recorded. When costs are considered, the inclusion of fine wine leads to a worsening of the Sharpe ratio in all cases. This result is a significantly more critical verdict on this diversification opportunity than was noted in the previous studies by Masset and Weisskopf (2010), Masset and Henderson (2010), Bouri (2014), Bouri et al. (2016) and Aytac et al. (2016). By contrast, our results confirm the studies which point out the high costs of investment in fine wine and which reach largely negative findings when analysing real investments in wine investment funds (Burton and Jacobsen, 2001, Masset and Weisskopf, 2015).

1 Introduction

Don't put all your eggs in one basket. One of the most important principles in investing is to hold a diversified portfolio. This means spreading one's capital between different investment forms in order to reduce dependency on one form of investment, because no matter how convinced one may be by an investment, things can turn out differently to how one imagined.

For centuries, investors have been practicing the concept of spreading financial investments widely. By investing in securities of different geographical origins and from various issuers, the risk of a portfolio is reduced. The fundamental principles of portfolio diversification can be seen in the brochures of investment funds as early as 1870 (Rutterford and Sotiropoulus, 2016). The theoretical basis of portfolio theory was laid down by Harry Markowitz (Markowitz, 1952). Via the diversification of assets, the risk of a portfolio (as measured by the standard deviation) can be reduced while maintaining the same expected return, or the returns can be increased without also increasing the risk involved. In order to achieve this, one has to not only take the individual risk into account when selecting equities, but also the price trends of various forms of investments over the course of time. When one asset loses value, another can gain. So, if they are not completely correlated, portfolio risk can be reduced by mixing different forms of investment. Spreading one's investments is possible in various ways. It can be done via different asset classes, or also by investing internationally in a range of countries and currencies. The best-known asset classes are equities, bonds and money-market instruments. These are referred to as classical or traditional asset classes, while all other types are described as alternative asset classes.[1] The price movements of alternative asset classes are different from those of the classical asset classes, which is what makes them so interesting. This is also the reason why they are suitable for diversification and for reducing portfolio risk.

In the aftermath of the global financial crisis in the years 2008/2009, interest rates fell significantly worldwide. Due to the very low returns, the prices of many bonds were noticeably higher than their redemption value. Until the point in time when the bond is repaid, there will therefore be price losses, so other investment opportunities with a chance of price increases thus gain in attractiveness. This is also one of the reasons why there has been impressive growth in alternative investments (Kräussl, 2017). Alongside hedge funds, private equity and commodities, so-called exotics or collectables are also increasingly being viewed as investment opportunities.[2]

An issue which has been debated more frequently in recent years is investment in wine (Ashton, 2010), whereby the positive aspects of diversification are particularly emphasized. The foundation for this is the fact that the factors which influence wine prices differ significantly from those which influence the prices of classical asset classes. The main influences on wine prices are the weather and thus the quality of the wine, the origin and classification[3] of the wine, and their evaluation by independent experts (above all Robert Parker).[4] The great influence the weather has is indisputable: a lot of rain in winter, a growth period with warm temperatures and a dry phase towards the end of the growth period and during harvesting are

---

[1] There is no standardised definition of the term 'alternative investments'. Everything which is not stocks or bonds should therefore be viewed as alternative investments (Fraser-Sampson, 2011).

[2] An overview of studies on the performance of collectables can be found in Burton (1999).

[3] The classification of wines in Bordeaux was first carried out at the time of the 1855 World Exhibition for wines from Medoc. The first growths (at that time there were four, since 1973 there have been five) were classed as Premier Grand Crus. These are the Chateaux Haut-Brion, Lafite-Rothschild, Latour, Margaux and Mouton-Rothschild.

[4] See Ashenfelter et al., 1995, Ali et al., 2010, Dubois and Nauges, 2010, Hay, 2010, Masset et al., 2015, Ashton, 2016.

ideal conditions for high-quality wines. The influence of expert evaluations is more limited, because the quality of the wine is initially determined by the weather and the know-how of the producer. The assessment by experts ultimately only confirms the quality determined by the above-mentioned factors (Lecocq, 2006). In addition, an IMF study also established economic growth in the emerging markets as an important influence on the development of wine prices (Cevik, 2011). The factors which influence wine prices thus differ considerably from the factors which determine the values of traditional asset classes. This is already a good indicator that wine could be suited towards improving the risk-return profile of securities portfolios.

For wine to be suitable as an investment, certain conditions must be fulfilled. The term fine wine investments is used when the following are in place: (1) There is a secondary market, (2) the wine has the potential to improve in the bottle, (3) the winery has a longer track record in the production of high-quality wines, and (4) there are positive evaluations by independent experts. The annual trading volume in the fine wine market is estimated to be around $4-5bn (in 2004 this was only approx. $ 1bn)[5]. Unlike other commodities such as gold or oil, fine wine is not traded on a standardized basis; there are too many different wines and too many trading places. The founding of the London International Vintners Exchange (Liv-ex) in 1999 was therefore a milestone in the professional trade with fine wine. Liv-ex is the world's most important electronic exchange for fine wine. It provides a weekday market for wine dealers, traders and brokers. The difference to auction houses is that only registered traders are allowed to trade here. In addition, the wines traded here are solely those which comply with the definition of fine wine, and not antique wines. Liv-ex calculates and publishes its own research and price index and follows strict transparency rules. The daily database of Liv-ex goes back as far as 2004 and thus provides sufficient history for the analysis in this study.[6]

Studies have been dealing with wine as an investment since the end of the 1970s. Krasker (1979) concluded that the returns from storing wine are no higher than from investing in no-risk investments. His database were the prices of wines from Bordeaux and California which were auctioned between 1973 and 1977 by Heublein, at the time a very well-known wine trader. Jaeger (1981) observed a longer period from 1969 to 1977 and reduced the assumed storage costs by using Freemark Abbey Winery as a basis for his calculations. He determined a performance above risk-free investments. In a long-term study covering the period 1900 to 2012, Dimson et al. (2015) came - on the basis of auction results and a 'value-weighted arithmetic repeat sales regression'[7] approach - to a geometrically-weighted average return of 5.3%. Taking storage and insurance costs into account, this is reduced to 4.1% (transaction costs can reduce the result further). For their study, the authors collected the prices of five wines. The wines are: Haut-Brion, Lafite-Rothschild, Latour, Margaux and Mouton-Rothschild. The majority of the prices originate from Christie's auction house in London. For an overview of the studies dealing with the issue of the performance of wine investments, see Storchmann (2012).

For this paper, studies which mainly focus on the diversification features of wine are of great relevance. These studies use time series of indices as their basis for calculations, and the majority of them come to positive results. According to them, the risk-return profile is significantly improved by including wine in a portfolio of traditional asset classes. In a study on the

---

[5] Most of the trading volume of wine is accounted for by dealers, and approx. 10% by auctions of the Liv-ex wine exchange (source: liv-ex.com).
[6] Bouri (2013) argued the case for using Liv-ex index data: it can be viewed as a reference for the fine wine market.
[7] Price trends are calculated here on the basis of transactions for exactly the same wines over time.

period from 1983 to 2002, Kumar (2010) compared the Fine Wine 50 index[8] with the FTSE 100 (UK equities), the Dow Jones Industrial Average (US equities), UK treasury bonds and US 30-year treasury bonds. His results show that investing in wine is more profitable than investing in stocks or bonds. Its volatility is higher than that of bonds, but lower than that of equities. The positive diversification effects in particular are emphasized. The portfolio with the highest return in Sterling and US dollars consisted entirely of wine. The minimum variance portfolio in Sterling consisted of 18% wine, 4% equities and 78% bonds. In US dollar it consisted of 1% wine, 1% equities and 98% bonds. It was shown that the Sharpe ratio[9] of the portfolio was higher with the inclusion of wine than without it.

In an analysis covering the period 1996 to 2007, Masset and Henderson (2010) showed on the basis of data from the Chicago Wine Company[10] that investing in wine is advantageous. An attractive return was obtained and the correlation with equities was not significant. This means that investment in wine leads to a meaningful diversification of portfolios. The portfolio with the highest Sharpe ratio contains a higher proportion of wine than equities.

Masset and Weisskopf (2010) established that for the period 1996-2009 that the General Wine Index[11] and above all the Grand Crus from Bordeaux from the best vintages achieved a better return than stocks, and at the same time exhibited lower volatility. A look at other wine indices also confirms that higher returns and lower risks were obtained than in the Russell 3000 Index.

In another study, Bouri (2014) compared two portfolios over the period 1988-2013 against a benchmark portfolio of 100% US stocks (S&P 500). The first of these portfolios contained 50% fine wine (Liv-ex-Fine Wine Investables Index)[12] and 50% US stocks. The second portfolio was a minimum variance portfolio consisting of US stocks and fine wine. It comes to the conclusion that adding wine investments to an equities portfolio leads to a reduction of risk. Particularly in times in which equities are losing value, wine shows its positive diversification effects.

Aytac et al. (2016) considered the period 2007-2014 in their study. On the basis of Liv-ex indices[13] as well as WineDex indices[14], gold, French equities (CAC indices), bonds (EMTX)[15] and a worldwide equities portfolio (MSCI World), they came to the conclusion that the inclusion of Bordeaux wines improved the efficiency in every case. Investors thus obtain a higher return for the same risk if they include wine in their portfolios. The results for the WineDex indices are clearly better than those of the London Liv-ex wine exchange. Both the Sharpe ratios and the modified Sharpe ratios[16] improved.

---

[8] This is an index which was calculated on the basis of auction results. It contained wines from ten producers from five vintages. The vintages were from 1961-1990.

[9] The Sharpe ratio, also called the reward to variability ratio, measures the average return earned in excess of a risk-free interest rate per unit of volatility.

[10] The Chicago Wine Company (TCWC) was founded in 1974 and held its first wine auction in April 1977. It was thus the second company after Heublein to hold wine auctions in the USA.

[11] The index data is from the Chicago Wine Company and contains all auctions which took place from January 1996 to February 2007.

[12] The index of the Liv-ex fine wine market in London consists of Bordeaux wines from the 24 leading producers. The selection takes place on the basis of evaluations by the wine critic Robert Parker (at least 95 out of 100 points).

[13] The indices are entirely transparent and can be viewed on the web page https://www.liv-ex.com/news-and-insights/indices/.

[14] The indices are calculated on the basis of wine auctions in France. They consist of 40 representative wines of the region from the last ten vintages. No information is available on the weighting of the index components.

[15] The EMTX indices cover the whole of Europe and were developed in 2003 by EuroMTS and Euronext.

[16] The modified Sharpe ratio assumes a skewed rather than a normal distribution of returns.

Bouri et al. (2016) analyzed monthly data for the period 2001-2014 and focused on the advantages in terms of diversification of investing in wine at times in which traditional asset classes are doing poorly: they examined the function of wine as a hedge and safe haven. For investments in wine, the Liv-ex Fine Wine Investables Index was used. The UK stock market was represented by the FTSE 100 Index. In the periods indicated, a differentiation was made between bull markets (from 2004-2007 and 2009-2014) and bear markets. They came to the conclusion that the risk-return ratio of a portfolio is improved by adding wine to it. Fine wine is thus considered to function as a hedge in negative market periods.

The main issue in this study is whether fine wine as an alternative asset can make a positive contribution towards diversification in a portfolio which is already broadly diversified in terms of the traditional assets of stocks and bonds. A positive contribution towards diversification is an increase of the annualized return over the observation period with the same level of risk (as measured by standard deviation) or a reduction of risk with the same return. In order to consider return and risk, the Sharpe ratio is used. The Sharpe ratio measures the over-return (the return in excess of the risk-free interest rate) per unit of risk. The higher the Sharpe ratio, the more attractive is the portfolio. In the studies mentioned, the analysis is carried out on the basis of index data. Index figures do not contain any costs and are merely useful for a theoretical view, so in this study calculations on the basis of real investments are used in addition to calculations on the basis of index values. A real investment can be carried out by purchasing investment funds. In the case of actively managed investment funds, investors delegate their investment decisions to experts who have specific expertise in this asset class. For this service, experts receive payment from their investors. In the case of fine wine, the delegation of investment decisions to experts appears particularly meaningful. This is because (1) there is no central market place and therefore there is no single price, (2) it is difficult to obtain access to market-relevant information due to a lack of regulation, and (3) high costs can be mitigated by economies of scale. In order to measure the success of funds which invest in fine wine, Masset and Weisskopf (2015) examined the performance of nine wine funds over the period 2000-2013. Only one of the nine funds considered [17] succeeded in achieving a higher Sharpe ratio than the benchmark index[18]. In their study, they also drew attention to the problem of liquidity in wine funds. In the years 2012 and 2013, for example, there were already two funds which had to close due to liquidity problems and contradictions in evaluations.[19] The reason for the closure of the funds was a high number of fund shares being returned by investors. If the contents of the fund cannot be liquidated quickly in such a case, the fund has to close.[20]

Alongside investing in actively managed investment funds, buying a passive index tracker is another way of investing in an asset class. Passive funds aim to replicate an index. Their performance is therefore very close to that of the underlying index. The advantage are the

---

[17] This was the Nobles Cru Fund. However, it should be noted that the fund used a method to evaluate the wines in its fund which it has developed itself and which is not comparable with other funds.

[18] The benchmark index is the Liv-ex-Fine Wine Investables Index.

[19] In autumn 2012 the Nobles Cru Fund had to close. After doubts about the correctness of the evaluation method of the fund wines, a large number of investors returned their shares. The Nobles Cru Fund had by far the best results in the Masset and Weisskopf study. In June 2013 The Vintage Wine Fund also had to close due to poor performance and accelerated sales.

[20] Some open real estate funds in Germany suffered the same fate in 2010. After doubts arose in the press about the evaluations of properties in the fund, shareholders rushed to return their holdings. When the liquidity of the fund had been used up it had to close. In some cases, it took years for the properties to be sold and for the investors to be reimbursed. In cases where shares in the fund were sold in a hurry, some investors suffered losses of up to a third of their investment.

significantly lower fees compared to actively managed funds. In this study, real investments are therefore considered on the basis of index funds. In the traditional asset classes, there are suitable funds available, but for investments in fine wine there has been no index fund until now. An index is therefore calculated which takes into account all of the actual costs of the investment.

In order to determine whether including fine wine in a portfolio is meaningful, the underlying investment universes have to be established. This applies to the stocks, bonds and the investment in fine wine. The goal is to ensure a well-diversified portfolio. Sufficient and easily realizable liquidity should also be present. This means that it must be possible to invest in the markets represented by the indices, and that they are not only available as a calculated index. On the equities side, the MSCI World Index was therefore chosen, which is broadly diversified according to countries as well as sectors of industry. It represents the global stock markets. The bond side is represented by the J.P. Morgan Global Government Bond Index: this global index is composed of the bonds of the leading industrial nations and is broadly diversified according to countries as well as maturities. With regard to the selection of a wine index, the majority of recent studies concentrate on the indices of the London International Vintners Exchange (Liv-ex). In the case of an index for investment in fine wine, transparency, investability and liquidity are decisive for the analysis. Transparency means that all information about the index must be available. This includes the index components, the weighting of the components, and the time and extent of changes in the index. The question of the price at which the wines in the index are calculated also has to be transparent. Investability means that all of the wines in the index are available, and that the wines are traded regularly. The criterion of liquidity means that the bid-ask spread is not too large, and that the order book consists of a sufficient number of positions. Due to the criteria to be fulfilled, the use of an index of the Liv-ex wine exchange in London is absolutely necessary, because only these indices fulfil all of the criteria. Which Liv-ex index is selected depends on the criterion of the highest level of liquidity. For years now, the five most sought-after wines at the London wine exchange have been the five first growths from Medoc/Bordeaux. The proportion of Bordeaux wines in total Liv-ex trading is falling, but they still accounted for over 70% in 2017. The Liv-ex Fine Wine 50 Index consists of the last ten physically available vintages of all of the Premier Grand Crus from Bordeaux. The composition of the index is changed at the end of June every year. The five wines from the oldest vintages leave the index, and the five newly-available wines are included in the index (for the current composition of the index, see Appendix 1).[21] At this point, one could argue that in the last few years in particular, the wines from Burgundy have increased in value particularly, and that one should perhaps choose an index containing these wines. However, the decisive criterion is liquidity, and for an investment in the index components the majority of the wines of the index, or even all of the wines, should be sufficiently tradeable. That only applies to the Liv-ex-50 Index.

For the comparison of a real investment, suitable index funds for the mentioned index universes have to be selected. To this end, an adequate history going back to March 2010 is

---

[21] Chateau Latour announced in 2012 that it was no longer going to sell any more wines *en primeur* (i.e. while they are still in the vats, as wine futures). In future, Latour will only put its wines on the market when it considers them ready to drink. For this reason, the last physically available vintage from Chateau Latour is 2011. For all other producers it is the 2015 vintage. The composition of the index with regard to Chateau Latour has had to adjust itself to the decision on the release of new vintages. This decision means that the 2011 vintage from Chateau Latour will remain in the index until a new vintage is sold. Until this point in time, the percentage proportion of the annual index adjustment (in relation to the entire index) will be reduced from 20% to 16% (8 out of 50 wines).

required, given that from that time the data for the calculation of the wine index including the costs is available. It is also important that at that point in time the index fund contained a certain amount of assets, thus signalizing sufficient tradability. A higher level of assets in the fund additionally ensures that the fix cost components of the fund do not have any significant negative effects on its performance. On the basis of these criteria, the iShares MSCI World Ucits ETF[22] (fund launched on 28 October 2005) was chosen for the investment in stocks. On the bonds side, it is the iShares Global Government Bond Ucits ETF (fund launched on 6 March 2009). As all of the costs are included in the price of the fund, the fund price can be used as the basis for the calculation of performance. For the real investment in fine wine, the Liv-ex 50 Index including costs is used. The data required for the calculation was made available by Liv-ex. Until now there have been no studies in which bid-ask spread data from the Liv-ex wine exchange has been taken into account. A particularly interesting question here is whether the costs of the actual investment in wine are higher than those of the investment in the traditional asset classes. Burton and Jacobsen (2001) already highlighted the high costs of an investment in wine. When transaction, insurance and storage costs are considered, the annual return is reduced by up to 3.7 percentage points according to their calculations.

Two portfolios were established as a starting point which can be considered as representative for different risk preferences. On the one hand, a portfolio is used which consists of 50% each of stocks and bonds (portfolio A1), and on the other hand a portfolio consisting of 25% stocks and 75% bonds (Portfolio B1). 10% of fine wine is added to both of these portfolios (Portfolio A2 and Portfolio B2).[23] In the calculation, an annual rebalancing[24] at the start of the year is presumed.

Due to the results of the previous studies, the first hypothesis is as follows:

Hypothesis 1: In a broadly diversified portfolio of traditional asset classes, a positive diversification effect is achieved by including fine wine.

The first null hypothesis to be examined is therefore:

Null hypothesis 1: In a broadly diversified portfolio of traditional asset classes, no positive diversification effect is achieved by including fine wine.

The results of Burton and Jacobsen (2001) give grounds for the presumption that the positive diversification effect will disappear when taking the costs of an investment in fine wine into account. The second hypothesis is therefore as follows:

Hypothesis 2: Taking all cost components into account, the risk-return ratio of a portfolio consisting of traditional asset classes is not improved by adding fine wine.

The second null hypothesis to be examined is therefore:

Null hypothesis 2: Taking all cost components into account, adding fine wine leads to an improvement of the risk-return ratio of a portfolio consisting of traditional asset classes.

---

[22] iShares is the name of the index provider. It belongs to the investment company BlackRock. UCITS is the abbreviation of Undertakings for Collective Investments in Transferable Securities and is a term which refers to the EU regulations on investment funds. ETF stands for Exchange Traded Fund and means that this is an exchange-traded index fund.

[23] The portfolios then consist of 45% stocks, 45% bonds, 10% wine, and 70% bonds, 20% stocks and 10% wine.

[24] The index weightings which have changed due to price trends are reset to their initial levels.

2 Data and methodology

2.1 Data

A calculation using historical data forms the basis of the analysis. The results of the calculations with historical data are considered to be an indication that they might also apply to the future. In a first step, a connection is made to the approaches used in the previous studies and a calculation is made using index figures (Aytac, 2016, Bouri, 2016). An index is composed of the price data of the index components and is normally calculated on a daily basis. It therefore represents a specific investment universe. Indices are benchmarks which supply a data basis for asset allocation and diversification decisions, as well as the calculation of key figures (Kumar, 2010). With regard to equities, the MSCI World Index was chosen for the calculation with index values. This global index is broadly diversified: It is composed of stocks from 23 countries, and the weighting is made according to market capitalization. Due to the size of the US stock market, the focus is therefore on US companies (index weighting approx. 60%, followed by Japan with around 9%). The bond market is represented by the J.P.Morgan Global Government Bond Index. This global index is composed of government bonds of the 13 most important industrial nations and is broadly diversified in terms of countries (the index weighting of the USA is approx. 40% here, followed by Japan with around 20%). The index covers the entire range of maturities from one to up to ten years to maturity. The Liv-ex Fine Wine 50 Index was chosen to represent fine wine. It consists of the last ten physically available vintages of all of the Premier Grand Crus from Bordeaux. The calculation on the basis of the indices covers the period from the beginning of 2004 to the end of May 2018. This period of time was chosen because it corresponds to the maximum availability of the data for the Liv-ex 50 Index of the London wine exchange. The calculation is carried out on the basis of monthly index figures which were taken from Datastream.[25]

As the indices are calculated on the basis of gross prices, costs are not taken into account. In a second step, the calculation on the basis of an actual investment is carried out. Investors who invest their own funds are predominantly interested in real opportunities and risks. A calculation on the basis of an index is therefore not very meaningful. This analysis is thus much closer to reality than previous studies. Stocks and bonds are both represented by investment funds. Here, the costs of an actual investment are taken into account. Passively managed index funds in particular are suited to reflect the exposure in a chosen asset class, because they are not subject to the risk of a deviation from the index in terms of returns which is due to active management. Deviations from the performance of the index are only due to the moderate fees which are included in index funds. For the calculation on the basis of the actual investments, the performance data of the iShares MSCI World Ucits ETF are used for the equities and that of the iShares Global Government Bond Ucits ETF for bonds. An adequate data basis is available for both of these stock exchange-traded index funds. The data required was made available by the fund company iShares. In the case of fine wine, on the other hand, there is no index fund available, and the calculation is made on the basis of an index where all costs are inclusive. The basis of the calculation is the Liv-ex 50 Index. The costs for the investment in the wines of the index via the Liv-ex trading platform were made available for the period from March 2010 to May 2018. This period thus forms the starting point for a comparison on the basis of an actual investment. Figure 1 summarizes the two perspectives.

---

[25] Information on the composition of the index is in the appendix.

Table 1: Basic data

	Period	Stocks	Bonds	Fine wine
Index comparison	01/2004-05/2018	MSCI World	JPM World	Liv-ex 50
Real investment	03/2010-05/2018	iShares MSCI World Ucits ETF	iShares Global-Government Bond Ucits ETF	Liv-ex 50 including costs

2.2 Methodology

In order to analyze the diversification properties of fine wine, the above-mentioned portfolio components must be combined into an overall portfolio. The composition of the portfolio is determined by the degree of risk aversion of the subjects. The higher the degree of risk aversion, the lower the proportion of the riskier asset. We oriented ourselves towards the system used in previous studies (Canner et al. 1997, Masset and Weisskopf, 2010, Aytac, 2015), which linked the proportion of equities in the portfolio with the willingness to take risks. We can use two portfolios as an example: they only differ in the size of the proportion of stocks and bonds in the portfolio. Portfolio A1 consists of 50% stocks and 50% bonds. Portfolio B1 consists of 75% bonds and 25% equities. Alternative investments or exotic asset classes are usually only included in small portions, as they are subject to special risks. Studies which have analyzed the inclusion of alternative asset classes used between 5% and 15% (Bessler and Wolff, 2015). We followed this system and considered the inclusion of 10% fine wine to be appropriate. Portfolio A2 is thus composed of 45% stocks, 45% bonds and 10% fine wine. After the inclusion of fine wine, Portfolio B2 is structured as follows: 20% stocks, 70% bonds and 10% fine wine. Figure 2 summarizes the portfolios.

Table 2: Portfolio compositions

	Portfolio A1	Portfolio B1	Portfolio A2	Portfolio B2
Stocks	50%	25%	45%	20%
Bonds	50%	75%	45%	70%
Fine wine	0%	0%	10%	10%

In the analysis, the portfolio is composed of up to three parts. Due to increases and decreases in values, the index weightings shift over the years. Indices which have a better performance than others present in the portfolio obtain a higher index weighting. In order to compensate for the shifting of the index weightings, rebalancing is carried out at the beginning of the year. The index weightings are thus reset to the initial level.

In order to judge performance, the annualized return of the respective portfolio is calculated over the observation period. For risk, the standard deviation is calculated. The standard deviation represents the overall risk of the portfolio; however, the simultaneous consideration of return and risk is decisive. The Sharpe ratio, also called the reward to variability ratio, is a suitable tool for this purpose. It measures the average return earned in excess of the risk-free

rate per unit of volatility or total risk. The higher the figure, the better. The Sharpe ratio is calculated according to the following formula (Sharpe, 1964):

$S(x) = (rx - Rf) / StdDev\ (x)$,

whereby $S(x)$ is the Sharpe ratio, rx is the portfolio return, Rf the risk-free interest rate, and StdDev (x) is the standard deviation. In US dollars, the 3-Month US Treasury Bill rate[26] was used as the risk-free interest rate (Rf). In euros it is the 3-month Libor rate[27].

The calculation was carried out in the base currency of US dollars as well as in euros. The dollar is the world's most significant investment currency, with the highest trading volume of any currency. It can thus also be viewed as the most representative currency. The euro is the world's second most-important reserve currency. The data series of the stock and bond index are available in US dollars as well as euros. The wine index used is quoted in sterling and is changed into US dollars[28] and euros. Table 3 shows an overview of all the portfolios which were compared.

For the calculation on the basis of the actual investments, the performance data of the iShares MSCI World Ucits ETF are used for equities and that of the iShares Global Government Bond Ucits ETF for bonds. There is no index fund for the investment in the Liv-ex 50 wine index. The calculation of the Liv-ex 50 is carried out on the basis of traded prices or the mid-prices[29] of the index components. As is usual in index calculations, transactions fees, bid-ask spreads, and storage and insurance costs are not taken into account. Transaction costs are the fees which are charged for buying and selling. At the Liv-ex wine exchange they are a percentage of the transaction volume. Bid-ask spreads are the difference between the best buying price and the best selling price on the stock exchange. As the Liv-ex 50 wine index is calculated from the mid-prices between the bid and ask price, there is a price deviation when transactions are made. This price difference has to be taken into account in the calculation of the index including costs. The following costs are therefore included in the calculation of the index including costs: (1) The cost of the initial investment[30], (2) the cost of the annual change of the composition of the index[31], (3) the annual storage and insurance costs for the storage of the wines in the wine exchange's warehouse, and (4) the annual subscription fees for access to the stock exchange. An exact list of the costs can be viewed in Appendix 2[32]. An annual management fee, such as those contained in stock and bond funds, is ignored. The bid-ask daily data for the

---

[26] The data was taken from the website of the Federal Reserve Bank in St. Louis, https://fred.stlouisfed.org. For the period 01/2004 to 05/2018 the average interest rate was set at 1.22%, and for the period 03/2010 to 05/2018 it was set at 0.3%.

[27] The data is available from the website https://de.global-rates.com. For the period 01/2004 to 05/2018 the average interest rate was set at 1.36%, and for the period 03/2010 to 05/2018 it was set at 0.25%.

[28] The rate of the US dollar to the British pound is referred to as the cable. The reason for this is that a cable was laid on the bed of the Atlantic in the mid-19th century in order to enable messages to be transmitted.

[29] The mid-price method can be viewed at https://www.liv-ex.com/news-and-insights/indices using the link: Mid Price logic here.

[30] One-off purchase of the portfolio at the ask price plus transaction costs.

[31] The five oldest vintages are sold, and the five new physically available vintages are bought (spread plus transaction costs, purchase at the ask price, sale at the bid price). The changeover of the vintages of Chateau Latour takes place on an irregular basis and is dependent on the release of new vintages.

[32] As some fees are volume-dependent, an investment amount of £10 million is assumed.

individual wines of the Liv-ex 50 Index are used to calculate the costs of an initial investment and the costs of the annual index adjustments.[33]

Table 3: Portfolio comparisons, hypotheses

Index calculation	Hypothesis 1: Portfolios with a higher Sharpe ratio
Portfolio A1 in USD vs. Portfolio A2 in USD	A2
Portfolio B1 in USD vs. Portfolio B2 in USD	B2
Portfolio A1 in EUR vs. Portfolio A2 in EUR	A2
Portfolio B1 in EUR vs. Portfolio B2 in EUR	B2
Real investment	Hypothesis 2: Portfolios with a higher Sharpe ratio
Portfolio A1 in USD vs. Portfolio A2 in USD	A1
Portfolio B1 in USD vs. Portfolio B2 in USD	B1
Portfolio A1 in EUR vs. Portfolio A2 in EUR	A1
Portfolio B1 in EUR vs. Portfolio B2 in EUR	B1

3 Results

The index data for the period between the beginning of 2004 and the end of Mai 2018 in USD shows a surprising picture regarding performance and standard deviation as well as the Sharpe ratio (see Table 4). The annualized return of the wine index at +6.96% is below that of the equities (+7.13%) and above that of the bonds (+3.80%). Risk as measured by the standard deviation is highest in the wine index by a clear margin. Its figure of 29.07% is significantly above that of the equities (16.35%) and bonds (5.29%). The Sharpe ratios reflect this relationship between return and risk. The wine index has the lowest Sharpe ratio at 0.2. The bonds have the highest Sharpe ratio at 0.49, while that of the equities is 0.36. In order to calculate the risk-free interest rate, the average interest rate of the 3-month Treasury Bill rate of 1.22% was determined over the period.

The results do not, however, reveal anything about possibly positive characteristics in terms of portfolio diversification, because the deciding factor are the correlations. The portfolio with 50% stocks and 50% bonds (A1) achieved an annualized return of 6.04% with a standard deviation of 7.54%. Adding 10% fine wine[34] (A2) increases the return to 6.31% while simultaneously increasing the standard deviation to 9.07%. The Sharpe ratio shows a clear picture. The inclusion of the wine index reduces the Sharpe ratio from 0.64 to 0.56. It must be noted that this result speaks against including fine wine in this portfolio. The results of the second portfolio consisting of 25% stocks and 75% bonds (B1) are just as clear. The return of the portfolio consisting of traditional asset classes can be increased from 5.04% to 5.35% by including fine

---

[33] Given that the index is calculated according to the mid-price method, the transactions at bid or ask prices (at half the spread) are included in the calculation of the index including costs.

[34] The composition of the portfolio is then: 45% stocks, 45% bonds and 10% fine wine.

wine (B2). However, at the same time, the risk in terms of the standard deviation also rises clearly from 4.44% to 6.18%. The Sharpe ratio of 0.86 is reduced to 0.67 by the inclusion of fine wine, while the return-risk profile thus worsens considerably.

Expectations regarding the improvement of the risk-return profile by including fine wine were not fulfilled. On the contrary, the Sharpe ratios worsened in both cases. One could possibly make the selection of the Liv-ex 50 wine index responsible for this: the index consists entirely of Bordeaux wines. Since 2015, wines from Burgundy have achieved significantly higher returns than those from Bordeaux. However, this argument can be countered by the fact that the Liv-ex 50 was among the best-performing indices over the entire observation period, while wines from Burgundy only started to increase in value considerably from 2015 onwards. On the other hand, a look at liquidity speaks against this: a broader index with a larger number of less 'liquid' wines would clearly lose when the return including all costs is calculated. The advantage of the gross price increases in Burgundy wines would soon be counteracted by the high costs.

Table 4: Index comparison in US dollars (1/2004-5/2018)

Period 2004-5/2018	100% stocks (MSCI World in USD)	100% bonds (JPM World in USD)	100% Fine wine (Liv-ex 50 in USD)	Portfolio A1 (50% stocks, 50% bonds)	Portfolio A2 (45% stocks, 45% bonds, 10% fine wine)	Portfolio B1 (25% stocks, 75% bonds)	Portfolio B2 (20% stocks, 70% bonds, 10% fine wine)
Return (annualized)	7.13%	3.80%	6.96%	6.04%	6.31%	5.04%	5.35%
Standard deviation	16.35%	5.29%	29.07%	7.54%	9.07%	4.44%	6.18%
Sharpe ratio	0.36	0.49	0.20	0.64	0.56	0.86	0.67

As long as equities are represented by the globally-diversified MSCI World Index, bonds by the globally diversified JPM World Index and fine wine by the Liv-ex-50 Index, no positive diversification effect can be attached to the inclusion of fine wine in two differently-weighted portfolios. On this basis, its inclusion cannot be recommended. The diverging results of previous studies can be traced back to a different data basis: they used different time period, different indices and a different base currency.

In order to address the issue of the different base currency, we also carried out a calculation with the euro as the base currency (see Table 5). Here one can see a change in the overall picture in favor of fine wine. When Portfolio A1 with 50% stocks and 50% bonds is considered from a risk-return perspective, it has a Sharpe ratio of 0.2. This figure is identical to that of Portfolio A2 which includes 10% wine. In the case of Portfolio B2 consisting of 25% stocks and 75% bonds, the inclusion of 10% wine creates a slightly positive diversification effect. The

Sharpe ratio increased to 0.17 (previously 0.16) in the portfolio consisting of 20% stocks, 70% bonds and 10% wine.

Table 5: Index comparison in euros (1/2004 - 5/2018)

Period 2004-5/2018	100% stocks (MSCI World in EUR)	100% bonds (JPM World in EUR)	100% fine wine (Liv-ex 50 in EUR)	Portfolio A1 (50% stocks, 50% bonds)	Portfolio A2 (45% stocks, 50% bonds, 10% fine wine)	Portfolio B1 (25% stocks, 75% bonds)	Portfolio B2 (20% stocks, 70% bonds, 10% fine wine)
Return (annualized)	5.47%	2.19%	5.30%	4.40%	4.71%	3.41%	3.76%
Standard deviation	22.04%	12.03%	35.73%	15.12%	16.53%	12.79%	14.26%
Sharpe ratio	0.19	0.07	0.11	0.20	0.20	0.16	0.17

The index comparison can only provide a theoretical indication of diversification properties. The reason for this lies in the calculation of the index. Indices do not contain any costs, and it is therefore not possible to invest in them at the prices given. The cheapest way to invest in indices is via index funds. Index funds are

offered by a multitude of fund companies. They allow investors to invest in the index universe at low fees. Their performance is that of the index minus the costs. In the traditional asset classes, the availability of index funds is very high, and the costs are low. In the case of alternative asset classes, however, availability is very low, or non-existent as in the case of fine wine.

For the calculation of the Liv-ex 50 Index including costs, the spreads between the prices of the wines at which an immediate sale is possible (bid price) and the prices at which an immediate purchase is possible (ask price) play a decisive role. As the index is recomposed at the end of June every year, the buying and selling has to be done at exactly this point in time: wines which no longer form part of the index are sold at the current buying rate (bid price), and wines which are newly included in the index are bought at the selling rates (ask prices). Figure 1 shows the average spreads of all index figures over the period of analysis.[35]

Compared to the index data, the fund data for the period between March 2010 and May 2018 shows a strongly diverging picture regarding performance and standard deviation as well as the Sharpe ratio[36] (see Table 6). This is definitely also due to the changed time scale, but

---

[35] A glance at the Liv-ex trading page shows that during the changeover of the index the calculation of transactions at bid or ask prices is legitimate. A potential wine index fund with a fund volume of only £10 million would have to sell over forty 12 bottle cases of the current oldest vintage (2006) for the index changeover. Due to the need for a timely changeover, this would presumably only be possible at average prices below the bid price.
[36] The average risk-free interest rate was set at 0.3% for the period 03/2010-05/2018.

above all it is because this is now fund data which contains all the costs of an investment. The iShares MSCI World Ucits ETF USD exhibits an annualized performance of +9.69% with a standard deviation of 11.10%. The Sharpe ratio is 0.85. The iShares-FTSE-G7 Government Bond Ucits ETF USD only achieved an annualized return of +1.32% with a standard deviation of 5.69%. The Sharpe ratio is 0.18. The Liv-ex 50 Index including costs exhibits an annualized performance of 0.02% with a standard deviation of 17.42%. The Sharpe ratio is -0.02. Here one can already note the difference to the analysis of the index data: by looking at a real investment, all of the key figures are - as expected - worse. This is shown most clearly in the investment in fine wine.

Figure 1: Average bid-ask spreads of all index figures of the Liv-ex 50
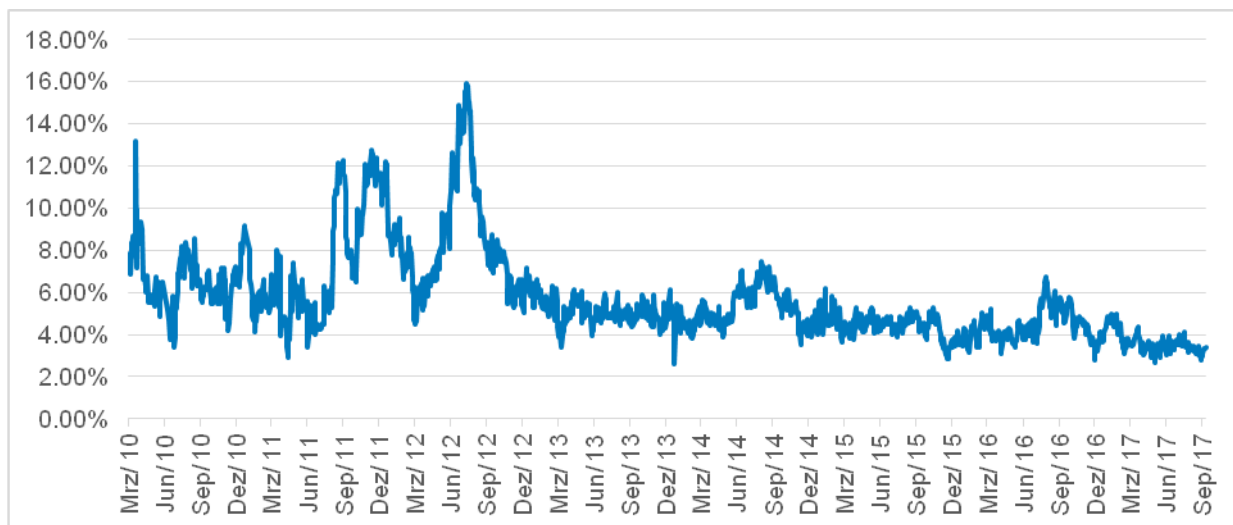
When the two portfolios including fine wine are compared, the annualized performance compared to the portfolios without fine wine is lower. For the portfolio with 50% stocks and 50% bonds, the annualized return over the period between March 2010 and May 2018 was 5.67% compared to 5.19% for the portfolio including 10% fine wine. At the same time, risk as shown by the standard deviation rose from 5.25% without wine to 5.91% with wine. The portfolio with 75% bonds and 25% stocks achieved a performance of 3.54% compared to only 3.06% for the portfolio with 10% wine. The standard deviation of the portfolio without wine was 4.18%. Including 10% fine wine increased the risk to a figure of 5.13% for the standard deviation. In both portfolios, the inclusion of 10% fine wine leads to an even clearer worsening of the risk-return profile than the calculation on an index basis. The Sharpe ratios also fall accordingly, from 1.02 to 0.83, and from 0.78 to 0.54.

Table 6: Index funds comparison in US dollars (3/2010 - 5/2018)

Period 3/2010-5/2018	100% stocks (MSCI World iShares ETF in USD)	100% bonds (FTSE G7 Governm. Bond iShares in USD)	100% fine wine (Liv-ex 50 in USD incl. costs)	Portfolio A1 (50% stocks, 50% bonds)	Portfolio A2 (45% stocks, 45% bonds, 10% fine wine)	Portfolio B1 (25% stocks, 75% bonds)	Portfolio B2 (20% stocks, 70% bonds, 10% fine wine)
Return (annualized)	9.69%	1.32%	0.03%	5.67%	5.19%	3.54%	3.06%
Standard deviation	11.10%	5.69%	17.42%	5.25%	5.91%	4.18%	5.13%
Sharpe ratio	0,85	0,18	-0,02	1,02	0,83	0,78	0,54

Here again, the funds are also considered in euros (see Table 7). The small diversification success of fine wine seen from the perspective of the index in euros cannot be repeated in a comparison of the index funds with the wine index in euros including costs. Both of the portfolios with the inclusion of 10% fine wine exhibit a poorer risk-return relationship than those without wine. The Sharpe ratios fall.

Table 7: Index funds comparison in euros (2/2010 - 5/2018)

Period 3/2010-5/2018	100% stocks (MSCI World iShares ETF in EUR)	100% bonds (FTSE G7 Governm. Bond iShares in EUR)	100% fine wine (Liv-ex 50 in EUR incl. costs)	Portfolio A1 (50% stocks, 50% bonds)	Portfolio A2 (45% stocks, 45% bonds, 10% fine wine)	Portfolio B1 (25% stocks, 75% bonds)	Portfolio B2 (20% stocks, 70% bonds, 10% fine wine)
Return (annualized)	7.69%	-0.51%	-1.79%	3.75%	3.28%	1.66%	1.19%
Standard deviation	17.27%	9.39%	20.70%	12.50%	12.87%	10.63%	11.06%
Sharpe ratio	0.43	n.a.[37]	n.a.	0.28	0.24	0.13	0.08

---

[37] The higher the Sharpe ratio of a portfolio the better is only valid for positive figures of the excess return (McLeod and van Vuuren, 2004). All calculated portfolios do have a positive excess return.

The results reflect the high cost burden involved in a real investment in fine wine. The transaction costs in connection with the high bid-ask spreads lead - in comparison to the traditional asset classes - to a clear worsening of the figures. The Total cost of ownership is below 0.28% (TER[38] 0.5%) for the equity funds, and below 0.17% (TER 0.2%) for the bond index funds. The annual total expense ratio in the case of the wine index (without taking the initial investment into account) is 1.15%, and with the initial investment it is 1.77%).   A management fee was not taken into account. [39] Studies which have not explicitly taken the aspect of costs into account cannot in good faith recommend an investment in fine wine in order to diversify a portfolio. Taking all cost components into account, the risk-return ratio of a portfolio consisting of traditional asset classes is worsened by adding fine wine. No positive diversification effects can be achieved.

The evaluation of the hypotheses takes place on the basis of a table of results (Table 8) in which the portfolios with the highest Sharpe ratios are shown. The portfolios with a higher proportion of equities are identified by the letter A, while those with a lower proportion of equities are identified by the letter B. The portfolios which do not include fine wine bear the number "1", while those which include 10% fine wine bear the number "2".

Table 8: Portfolio comparisons, results of the hypotheses

Index calculation	Results of hypothesis 1: Portfolios with a higher Sharpe ratio
Portfolio A1 in USD vs. Portfolio A2 in USD	A1
Portfolio B1 in USD vs. Portfolio B2 in USD	B1
Portfolio A1 in EUR vs. Portfolio A2 in EUR	A1=A2
Portfolio B1 in EUR vs. Portfolio B2 in EUR	B2
Real investment	Results of hypothesis 2: Portfolios with a higher Sharpe ratio
Portfolio A1 in USD vs. Portfolio A2 in USD	A1
Portfolio B1 in USD vs. Portfolio B2 in USD	B1
Portfolio A1 in EUR vs. Portfolio A2 in EUR	A1
Portfolio B1 in EUR vs. Portfolio B2 in EUR	B1

As can be seen from Table 3, hypothesis 1 must be discarded. The null hypothesis proved to be correct. Only in one of the four portfolios (Portfolio B2 in EUR consisting of 70% bonds, 20% stocks and 10% fine wine) does a slight increase of the Sharpe ratio occur due to the inclusion of fine wine. In the other three cases, the inclusion of wine does not lead to an improvement

---

[38] In the case of both index funds, the total cost of ownership is below the total expense ratio (TER) shown. The TER includes all annual administration fees. Transaction and trading fees are not taken into account. At the same time, returns from securities lending are also not included. These returns exceed other fees, so that the total cost of ownership (calculated as the difference between the performance of the index and the index fund's performance) is lower than the TER.

[39] Nevertheless, these costs are clearly below the 3.7% p.a. calculated by Burton and Jacobsen (2001). Due to the establishment of the Liv-ex wine market, investment in fine wine has become considerably cheaper.

of the Sharpe ratio. When a real investment including costs is considered, hypothesis 2 is confirmed. Null hypothesis 2 has to be discarded. All of the portfolios in US dollars and euros have a higher Sharpe ratio without fine wine. No positive diversification effects can be established. [40] This study thus assesses the possibilities for diversification much more critically than the previous studies by Masset and Weisskopf (2010), Masset and Henderson (2010), Bouri (2014), Bouri et al. (2016) and Aytac et al. (2016). At the same time, those studies are confirmed which highlight the high costs involved in fine wine investment, and which come to predominantly negative conclusions about real investments in wine investment funds (Burton and Jacobsen, 2001, Masset and Weisskopf, 2015).

4 Summary

In the search for alternative asset classes for portfolio diversification, the positive qualities of investment in fine wine have increasingly been reported on in the past 15 years. Whereas older studies had to fall back on data from wine auction houses, newer ones predominantly use index data from the London fine wine market Liv-ex, which was established in 1999.

However, making calculations on the basis of index data contains pitfalls. Index data differs from a real investment in that it does not take into account the costs which are incurred. This question is of particular relevance for investments in fine wine, as there have been no such index funds until now. Some wine investment funds do exist, but their strategies are neither transparent nor is the performance convincing (Masset and Weisskopf, 2015).

This study took into account the possible costs of a real investment in all of the asset classes considered. With this data, the diversification features of the inclusion of fine wine in a global portfolio which was already broadly diversified were examined and compared with an analysis on the basis of index data alone.

At an index level, the inclusion of fine wine merely leads to a slight improvement of the annualized return, but at the same time to an increase in risk. Only in the case of one portfolio in euros was the Sharpe ratio somewhat better after the inclusion of 10% fine wine. When considering an actual investment, the high costs of investing in fine wine play a role. In US dollars and euros, the annualized returns are lower in comparison to the portfolios without fine wine, and at the same time the risk levels are higher. The inclusion of wine always leads to a decrease in the Sharpe ratio.

The conclusion on the diversification properties of fine wine in a portfolio which is already broadly diversified in traditional asset classes can be - fittingly - described as sobering. Fine wine does not lead to any improvement of the Sharpe ratio from the index perspective in US dollars, and when the base currency is euros only a slight increase is achieved in one out of two portfolios. Viewed according to costs, there is not a single case where the risk-return profile was improved by adding fine wine. This finding is clearly contradictory to those of the previous studies. However, on the positive side it can be noted that the costs of an actual investment in fine wine have been noticeably reduced by the establishment of the Liv-ex Wine Exchange.

---

[40] A calculation with a uniform moderate spread of 4% which diverges from the actual bid/ask spreads as could be observed in recent years does not lead to any diverging results with regard to the diversification properties of fine wine.

Acknowledgments

References

Ali, H. H. / Lecocq, S. / Vissei, M. (2010): The Impact of Gurus: Parker Grades and en primeur Wine Prices*. Journal of Wine Economics, 5(1), 22-39.

Ali, H. H. / Nauges, C. (2007): The Pricing of Experience Goods: The Example of En primeur Wine. American Journal of Agricultural Economics, 89(1), 91-103.

Ashenfelter, O. / Ashmore, D. / Lalonde, R. (1995): Bordeaux Wine vintage quality and the Weather. Chance, 8, 7-14.

Ashton, R. H. (2016): The Value of Expert Opinion in the Pricing of Bordeaux Wine Futures, Journal of Wine Economics, 11(2), 2016, 261–288.

Aytac, B. / Hoang, T.-H.-V. / Mandou, C. (2016): Wine: To drink or to invest in? A study of wine as an investment asset in French portfolios, Research in International Business and Finance, 36, 591-614.

Aytac, B. / Hoang, T.-H.-V. / Mandou, C. (2014): Le management du vin, entre terroirs et mondialisation, une filière en pleine mutation. De Boeck, Brussels, 150 pp.

Bessler, W. / Wolff, D. (2015): Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies. Journal of Banking & Finance, 60, 1-20.

Bouri, E. I. (2013): Do Fine Wines Blend with Crude Oil? Seizing the Transmission of Mean and Volatility Between Two Commodity Prices. Journal of Wine Economics, 8(1), 49–68.

Bouri, E. I. (2014): Beyond the negative relation between return and conditional volatility in the wine market, International Journal of Wine Business Research, 26(4), 279-294.

Bouri, E. I. / Roubaud, D. (2016): Fine Wine and Stocks from the Perspective of UK Investors: Hedge or Safe Haven? Journal of Wine Economics, 11(2), 233-248.

Burton, B. J. / Jacobsen, J. B. (1999): Measuring Returns of Investments in Collectables. Journal of Economic Perspectives, 13(4), 193-212.

Burton, B. J. / Jacobsen, J. B. (2001): The Rate of Return on Investment in Wine. Economic Inquiry, 39(3), 337-350.

Canner, N. / Mankiw, N. G. / Weil, D. N. (1997): An Asset Allocation Puzzle. The American Economic Review, Vol. 87 (1), 181-191.

Cevik, S. / Sedik, T. S. (2011): A Barrel of Oil or a Bottle of Wine: How Do Global Growth Dynamics Affect Commodity Prices? IMF working paper.

Dimson, E. / Rousseau, P. L. / SPAENJERS, C. (2015): The price of wine, Journal of Financial Economics, 118, 431-449.

Dubois, P. / Nauges, C. (2010): Identifying the effect of unobserved quality and expert reviews in the pricing of experience goods: Empirical application on Bordeaux wine. International Journal of Industrial Organization, 28, 205-212.

Fraser-Sampson, G. (2011): Alternative Assets. Investments for a Post-Crisis world. Wiley.

Gecy, C. (2016): The New Diversification: Open your Eyes to Alternatives. The Journal of Private Equity, 20(1), 72-81.

Hay, C. (2010): The political economy of price and status formation in the Bordeaux en primeur market: the role of wine critics as rating agencies, Socio-Economic Review, 8, 685–707.

Jaeger, E. (1981): To Save or Savor: The Rate of Return to Storing Wine. Journal of Political Economy, 89(3), 584-592.

Kräussl, R. (2017): The search for yield: Implications to alternative investments. Journal of Empirical Finance, 44, 227-236.

Krasker, W. S. (1979): The Rate of Return to Storing Wines. Journal of Political Economy, 87(6), 1363-1367.

Kumar, M. (2010): Wine Investment for Portfolio Diversification. How Collecting Fine Wines Can Yield Greater Returns Than Stocks and Bonds. Wine Appreciation Guild, Libraries Australia, 179 pp.

Markowitz, H. (1952): Portfolio Selection, The Journal of Finance, 7(1), 77-91.

Markowitz, H. (1959): Portfolio Selection, Efficient Diversification of Investments, Cowles Foundation for Research in Economics at Yale University.

Markowitz, H. (1991): Foundations of Portfolio Theory, The Journal of Finance, 46(2), 469-477.

Masset, P. / Henderson, C. (2010): Wine as an Alternative Asset Class*. Journal of Wine Economics, 5(1), 887-118.

Masset, P. / Weisskopf, J.-P. (2015): Wine Funds: An Alternative Turning Sour? The Journal of Alternative Investments, 6-20.

Rutterford, J. / Sotiropoulus, D. P. (2016): Financial diversification before modern portfolio theory: UK financial advice documents in the late nineteenth and the beginning of the twentieth century. The European Journal of the History of Economic Thought, 23(6), 919-945.

Sharpe, W. (1964): Capital asset prices: a theory of market equilibrium under condition of risk. Journal of Finance, 19 (3), 425-442.

Storchmann, K. (2012): Wine Economics, Journal of Wine Economics, 7(1), 1–33.

Wood, D. / Anderson, K. (2006): What determines the Future Value of an Icon Wine? New Evidence from Australia. Journal of Wine Economics, 1(2), 141-161.

Appendix 1: Index compositions

Equities, composition of the MSCI World Index

Country	Index weighting	Number of companies
USA	60.26%	631
Japan	8.99%	321
United Kingdom	6.53%	102
France	3.98%	79
Germany	3.46%	65
Canada	3.45%	90
Switzerland	2.72%	37
Australia	2.46%	67
Hong Kong	1.35%	48
Holland	1.29%	21
Spain	1.06%	22
Sweden	0.93%	31
Italy	0.85%	24
Denmark	0.62%	17
Singapore	0.49%	26
Finland	0.38%	12
Belgium	0.37%	10
Norway	0.25%	9
Israel	0.18%	11
Ireland	0.18%	6
Austria	0.09%	5
New Zealand	0.06%	7
Portugal	0.06%	3

Bonds, composition of the JPM World Government Bond Index

Country	Index weighting
USA	40.25%
Japan	19.72%
France	8.17%
United Kingdom	7.15%
Italy	7.09%
Germany	5.46%
Spain	4.67%
Belgium	2.02%
Australia	1.68%
Holland	1.63%
Canada	1.33%
Denmark	0.50%
Sweden	0.31%
Term	
1-3 years	24.39%
3-5 years	19.28%
5-7 years	12.07%
7-10 years	12.56%
10+ years	31.69%

Fine Wine, Liv-ex 50 Index, equally weighted according to producer and vintage

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Chateau Haut-Brion				X	X	X	X	X	X	X	X	X	X
Chateau Margaux				X	X	X	X	X	X	X	X	X	X
Chateau Mouton Roth-schild				X	X	X	X	X	X	X	X	X	X
Chateau Lafite Roth-schild				X	X	X	X	X	X	X	X	X	X
Chateau Latour	X	X	X	X	X	X	X	X	X	X			

Appendix 2: List of costs of the Liv-ex 50 Index

An investment of £10 million (pounds sterling) is presumed.

Trading costs (bid-ask spread)

The Liv-ex 50 contains 50 wines. Every year at the end of June, the five oldest vintages are replaced by the five physically available new vintages. As the index is equally weighted, in this way 20% of the index is traded (currently 16%, until a new vintage of Chateau Latour is released). This index adjustment is carried out at the bid-ask spreads available at the end of June. The vintages to be sold are sold at bid prices, and the vintages to be bought are bought at ask prices. The first investment was made at the selling rates valid at that time (ask prices).

Transaction costs

1% transaction costs per transaction. It is assumed that a transaction volume of £20,000 per year is reached. This is the condition for a transaction fee of 1%.

Insurance

£10 per month plus 0.015%. This corresponds to 0.18% p.a.

Storage

£0.55 per case and month. From a volume of £10 million and at an average bottle price of £500, the price for a 12-bottle case is £6,000s. At an investment of £10 million, this would be 1,666 cases. 1,666 x 0.55 x 12 = 0.11% p.a.

Liv-ex subscription fee

Here, the Liv-ex 'gold' pricing package with direct market access is presumed. The fee amounts to £500 per month, £6000 per year. This is equivalent to 0.06% p.a.

Chapter 3

PORTFOLIO DIVERSIFICATION:
THE INFLUENCE OF HERDING, STATUS-QUO BIAS, AND THE GAMBLER'S FALLACY

with Ibrahim Filiz, Markus Spiwoks and Kilian Bizer
contribution Thomas Nahmer: 35%

Abstract

This experimental study examines the influence of herding (following the majority of fellow gamblers or the most successful gambler (guru)), status-quo bias, and the gambler's fallacy on diversification behavior. We find that neither herding nor status-quo bias contributes significantly to non-optimal portfolio choices. The gambler's fallacy, however, plays an important role in these decisions. Many subjects appear to find patterns in a history of random events and then use these "patterns" to infer the sequence of future events. The gambler's fallacy is significantly responsible for the fact that the optimal structure of a portfolio is considered in only 37.7% of all choices made by an investor.

1 Introduction

Markowitz (1952) shows that it is useful for risk-averse investors to split capital among different investment instruments. Practice shows, however, that investors often have strongly underdiversified portfolios.[1]

Experimental economic research increasingly addresses the question of why investors seem to find it so difficult to make useful portfolio diversifications. Gubaydullina and Spiwoks (2015) show that many investors have difficulty dealing with the correlations of income return developments.[2] The meaning of the correlations is systematically misjudged. Take the example of 1/n heuristics, where investors distribute their capital equally among all investment alternatives seemingly without noticing or caring how strongly the income returns of these instruments are correlated. Morrin et al. (2012) provide evidence that many subjects tend toward 1/n heuristics (for similar findings, see Fernandes, 2013; Baltussen and Post, 2011). Rieger (2012) reveals that investors systematically miscalculate the probabilities of occurrence. Fellner, Güth, and Maciejovsky (2004) conclude that investors often suffer from an illusion of expertise, and hence overestimate the advantageousness of their own choice of investment. Choi, Laibson, and Madrian (2009) note that diversification decisions are distorted by the phenomenon of mental accounting. Weber, Siebenmorgen, and Weber (2005) detect that investors are subject to a home bias when choosing investment instruments for a portfolio.

We conducted expert discussions with high-ranking bank managers,[3] which revealed other possible reasons for suboptimal diversification decisions. It seems possible that many subjects are distracted from optimal diversification by observing the investment choices of other investors (herding). Being influenced by the successful investment decisions of prominent investors can play a significant role in this process (guru effect). Furthermore, optimal diversification can be hindered by subjects holding on to existing portfolios (status-quo bias). Subjects can also be distracted from meaningful portfolio choices by exclusively following putative patterns of random events (the gambler's fallacy).

The present study therefore addresses the question of whether or not (1) herding, (2) status-quo bias, and (3) the gambler's fallacy do, indeed, sidetrack subjects from making optimum diversification decisions. The literature provides multiple indications of the significant influence that these phenomena can exert on economic decisions.

Looking at the herding literature first, the observation that subjects take their bearings from one another and thereby act as a herd traces as far back as Mackay (1841). Keynes (1936) points out the herding behavior of financial market actors and presents two possible explanations for it (reputational herding and investigative herding). Scharfstein and Stein (1990)

---

[1] See, e.g., Dimmock *et al.* (2016), Anderson (2013), Hibbert, Lawrence, and Prakash (2012), Goetzmann and Kumar (2008), Meulbroek (2005), Polkovnichenko (2005), Huberman and Sengmueller (2004), Agnew, Balduzzi, and Sundén (2003), Guiso, Haliassos, and Japelli (2002), Benartzi (2001), Benartzi and Thaler (2001), Barber and Odean (2000), Bode, van Echelpoel, and Sievi (1994), Blume and Friend (1975), and Lease, Lewellen, and Schlarbaum (1974).

[2] For similar results, see also Eyster and Weizsäcker (2011), Kallir and Sonsino (2009), and Hedesstrom, Svedsater, and Garling (2006).

[3] We thank Mr. Lothar Henning, Bethmann Bank Frankfurt, and Mr. Frank Weber, Sparkasse Lippstadt, for extensive talks concerning investment behavior of bank customers.

continue along these lines, adding fuel to a fierce debate that has been raging for the past 25 years. Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) show that herding can even occur when subjects behave rationally and make reasonable decisions (informational cascades). Devenow and Welch (1996) were the first to clearly differentiate between rational herding (reputational herding, investigative herding, and informational cascades) and irrational herding. There are numerous empirical findings that confirm herding behavior among actors on the financial market.[4] Therefore, it seems reasonable to consider herding as a possible origin of non-optimal portfolio diversification. To date, there have not been any experimental studies examining the potential influence of herding on diversification decisions.

A special case of herding is the guru effect. Gurus are highly ranked religious authorities in Hinduism and Buddhism. In Western cultures, the term "guru" also refers to leaders whose followers trust them blindly and uncritically, and the term "guru effect" has been used to describe the situation where private investors rigorously copy the decisions of prominent and very successful investors. The gurus' behavior is closely observed by many actors on the capital market, which is why it can lead to herding. In the research on this phenomenon, capital market simulation with interacting artificial agents (agent-based computational economics) has established itself as a reliable research method. The method has revealed that the network structure of communication among the agents significantly influences events on the capital market. Gurus are so-called super nodes that have numerous direct communication links with other capital market actors and, for this reason, can trigger herding.[5] Furthermore, the guru effect may contribute to distracting investors from optimal diversification decisions. There is as yet no research on how the influence of an investment guru impacts investor's portfolio decisions.

Many people find it difficult to make decisions and thus tend to avoid them, simply leaving things as they are, which is known as the status-quo bias (cf. Samuelson and Zeckhauser, 1988). The psychological processes of this behavior are explained in detail by Anderson (2003). Especially in situations when investors accede to an existing security portfolio (e.g., by inheritance), they often tend to postpone or even completely fail to adjust the portfolio structure. Even if different performances of the stocks in the portfolio lead to an unintended imbalance, many investors, out of dread of adjusting the portfolio, fail to take appropriate action. Aside from dread, this lack of action is often grounded in reluctance to take responsibility for the portfolio's future profit, or lack thereof. Many investors are afraid of regretting their own actions (cf. Inman and Zeelenberg, 2002; Zeelenberg et al., 2002; Kahnemann and Tversky, 1982). There are now some empirical findings on the status-quo bias as it applies to financial

---

[4] Huang, Wu, and Lin (2016), Choi (2016), Galariotis, Rong, and Spyrou (2015), Chang (2013), Kremer and Nautz (2013), Lin, Tsai, and Lung (2013), Belhoula and Naoui (2011), Boyson (2010), Kim and Jegadeesh (2010), Chiang and Zheng (2010), Spiwoks, Bizer, and Hein (2008), Chen, Wang, and Lin (2008), Walter and Weber (2006), Voronkova and Bohl (2005), Spiwoks (2004), Sias (2004), Ennis and Sebastian (2003), Chang, Cheng, and Khorana (2000), Nofsinger and Sias (1999), Wermers (1999), Choe, Kho, and Stulz (1999), Christie and Huang (1995), Lakonishok, Shleifer, and Vishny (1992), Klemkovsky (1977), Kraus and Stoll (1972).

[5] See, e.g., Panchenko, Gerasymchuk, and Pavlov (2013), Hein, Schwind, and Spiwoks (2008, 2012), Tedeschi, Iori, and Gallegati (2009, 2012), and Markose, Alentorn, and Krause (2004). Sumpter, Zabzina, and Nicolis (2012) show that a small number of leaders can heavily influence decisions.

market actors.[6] Numerous experimental studies also provide evidence of the status-quo bias (see, e.g., Geng, 2016; Yen and Chuang, 2008). Hence, it seems reasonable to consider the status-quo bias as a possible reason for non-optimal diversification. There is as yet only one experimental study that directly addresses this topic: Brown and Kagel (2009) yield information on the influence of the status-quo bias on non-optimal portfolio choices.

Looking now at the literature on the gambler's fallacy, we start with the experimental study by Gubaydullina and Spiwoks (2015) that found that irrelevant information can distract subjects from optimal diversification decisions. Considering the history of random events in evaluating random processes seems particularly tempting to many subjects, a phenomenon known as the gambler's fallacy. For example, if a coin toss shows "heads" three times in a row, many people assume that "tails" will show next. The history of unconnected random events, however, does not reveal anything about the future. The possibility for "heads" in the fourth toss is also exactly 50%. The gambler's fallacy has long been a subject of interest,[7] but no work has been done in the context of portfolio diversification.

We conduct an experiment to discover the reasons for insufficient portfolio diversification. In 15 periods, the subjects must make individual decisions about the structure of a portfolio of stocks. By taking a between-subjects approach, we examine the possible influence of herding, the status-quo bias, and the gambler's fallacy. We find that neither herding nor status-quo bias contribute significantly to non-optimal portfolio choices. The gambler's fallacy, however, plays an important role in these decisions.

2 Hypotheses and Experimental Design

2.1 Identification of Optimal Portfolios

Identifying optimal diversification decisions is difficult even in the easiest of cases, when there are only two stocks (A and B) involved that are independent in their income return development. Not only must the efficient frontier of all possible stock combinations be determined, but the investor's indifference curve must be considered (see Figure 1).

A strongly risk-averse investor (Subject 1) finds his ideal combination of stocks in the lower margin of the efficient frontier. A less risk-averse investor (Subject 2), however, finds his ideal combination of stocks in the upper margin of the efficient frontier. Although there are reliable empirical methods to differentiate between risk-averse, risk-neutral and risk-loving subjects,[8] capturing the exact layout of the indifference curves for a specific subject remains impossible.

---

[6] See, e.g., Freiburg and Grichnik (2013), Bryant, Evans, and Bishara (2012), Gubaydullina, Hein, and Spiwoks (2011), Kempf and Ruenzi (2006), Choi et al. (2004), Agnew, Balduzzi, and Sundén (2003), and Patel, Zeckhauser, and Hendricks (1991).

[7] See e.g. Chen, Moskowitz, and Shue (2016), Suetens, Galbo-Joergensen, and Tyran (2016), Stöckl et al. (2015), Powdthavee and Riyanto (2012), Barron and Leider (2010), Ayton and Fischer (2004), Clotfelter and Cook (1991), Tversky and Kahneman (1971, 1974).

[8] See, e.g., Lönnqvist *et al.* (2015), Charness, Gneezy, and Imas (2013), Crosetto and Filippin (2013), Dohmen *et al.* (2011), Eckel and Grossmann (2002, 2008), Lejuez *et al.* (2003), Holt and Laury (2002), and Gneezy and Potters (1997).

Some studies try to solve this problem by considering all stock combinations on the efficient frontier as an ideal choice. However, this approach does not consider that only one exact point of the efficient frontier can be deemed the optimal combination of stocks for an individual investor.

Figure 1: Identification of Optimal Stock Combinations in Consideration of the Efficient Frontier and Individual Risk Aversion, or Individual Indifference Curves
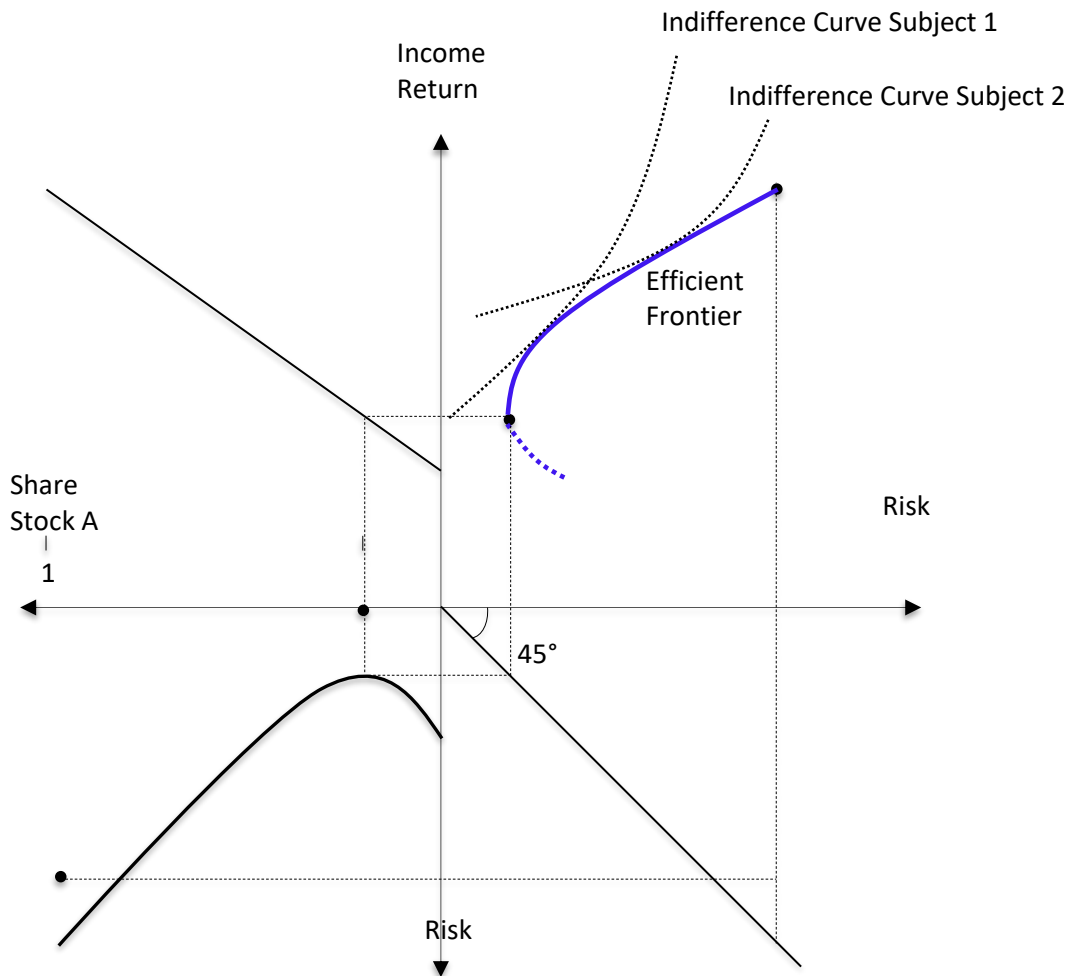
To avoid results so vague that interpretation of them is an exercise in futility, we employ the method of Gubaydullina and Spiwoks (2015): The subjects are offered two entirely uncorrelated alternatives for investment (A and B), which are identical regarding the expected income return and risk. By doing so, the efficient frontier is reduced to a single point (point Z in Figure 2). In this environment, it is of no importance whether a strongly or a less risk-averse subject makes the decision. In both cases, only the exactly equal mix of both investment alternatives (A and B) can be interpreted as the ideal combination of stocks. Only with this methodology can precise results in an experiment on diversification be obtained.

Figure 2: Precise Identification of an Ideal Combination of Stocks with a Punctiform Efficient Frontier (Point Z).
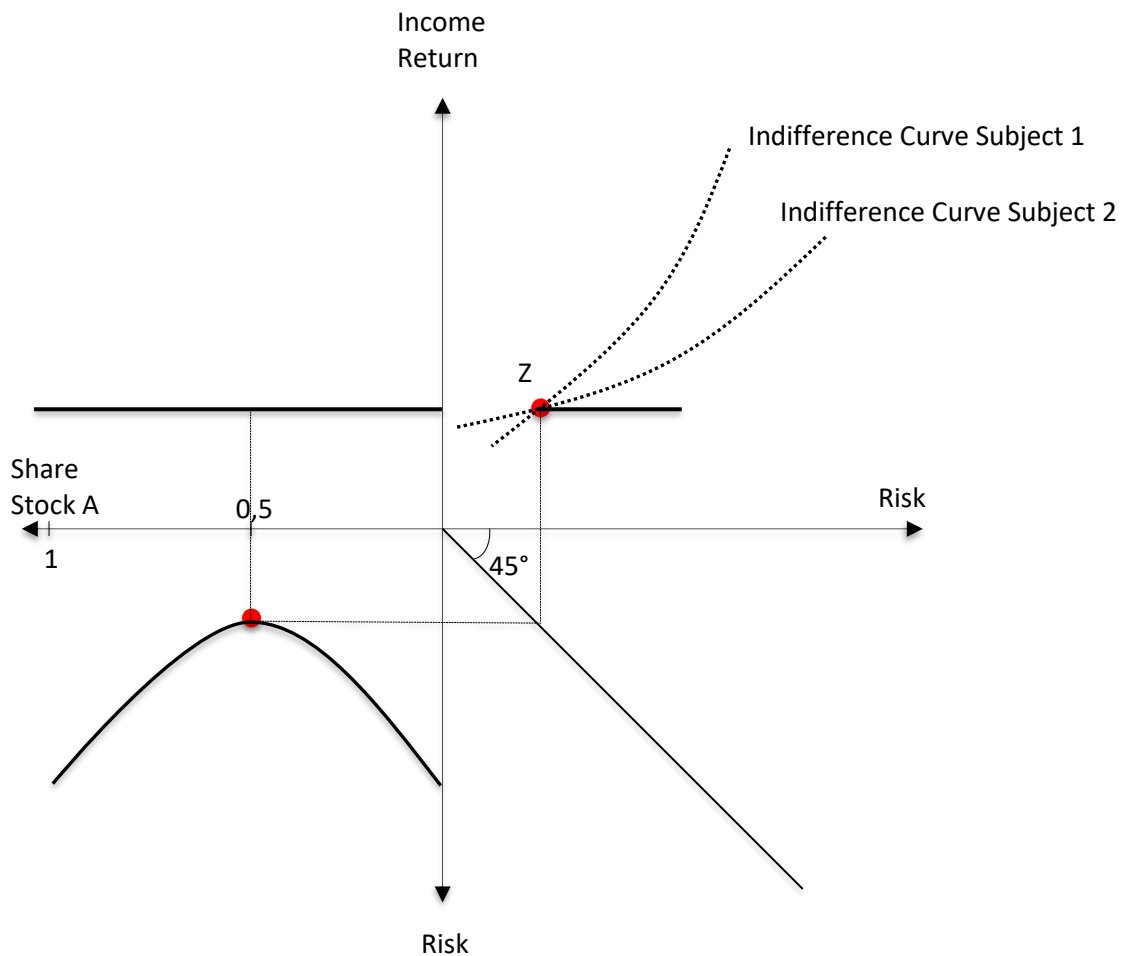
Ours is an individual decision experiment that follows the approach of Gubaydullina and Spiwoks (2015). The subjects can choose between two different risky securities, A and B. In each period, they assemble a four-stock portfolio. The possible portfolios are AAAA, AAAB, AABB, ABBB, and BBBB. Stock A and stock B both generate an income return of ±0 experimental currency units (ECU) or +7 ECU in each period. Both possible returns occur with a probability of 50% and follow a random process. Stock A's performance is independent of Stock B's performance. Both stocks thus have an expected value of 3.5 ECU. A portfolio consisting of four stocks is expected to generate a return of 14 ECU per period (see Table 1). The ECUs are converted into Euros in the ratio of 10:1, resulting in an expected value of €1.40 per period. Hence, the subjects can profit from their investment behavior if it is successful.

The five possible portfolios may have the same expected income return of 14 ECU per period but the exposure to risk – henceforth expressed by the standard deviation – is different for

each portfolio.[9] The standard deviation of the combination AAAA is 14.0, whereas the standard deviation of the combination AABB is only 9.9 (see Table 1).

Table 1: Expected Values and Standard Deviations of the Income Return for the Five Portfolios Considering the Possible Random Events for Stocks A and B in ECU

Random Events	A: +7; B: +7 (p_1 = 0.25)	A: +7; B: ±0 (p_2 = 0.25)	A: ±0; B: +7 (p_3 = 0.25)	A: ±0; B: ±0 (p_4 = 0.25)	E(r)	SD
AAAA	+28	+28	±0	±0	14	14.0
AAAB	+28	+21	+7	±0	14	11.1
AABB	+28	+14	+14	±0	14	9.9
ABBB	+28	+7	+21	±0	14	11.1
BBBB	+28	±0	+28	±0	14	14.0

p = probability of occurrence; E(r) = expected value of income return; SD = standard deviation.

2.2 Rational Strategy

A rational, risk-averse subject should always choose the combination AABB. Since the expected income returns of the five possible portfolios are identical, it is rational for each risk-averse subject to choose the portfolio with the minimum variance – independent of the degree of the subject's risk aversion.

This choice is intuitive. Regarding the structured components of the given stocks, the subjects can recognize the portfolio with the minimum variance without having to make any mathematical calculations. Using simple plausibility, it can be established that the income return level is most when both stocks A and B are equally represented in the portfolio (see Table 1).

However, considering the numerous empirical findings on the incapability or reluctance of subjects to make reasonable diversification decisions, we expect clear deviations from the rational strategy (always portfolio AABB) to occur in this experiment.

We thus arrive at our first hypothesis.

Hypothesis 1: The subjects are going to behave rationally, which means that they are going to exclusively choose the portfolio with the minimum variance (AABB).

The experiment consists of three treatments. The rational investment strategy (always AABB) is easily realizable in all three treatments. Therefore, the subjects' tendency toward rational diversification decisions is analyzed in all three treatments (see Table 2).

---

[9] Whether or not the subjects exhibit herding behavior, whether or not they are subject to the status-quo bias, and regardless of whether or not they fall prey to the gambler's fallacy does not influence their payout. If herding, status-quo bias, or the gambler's fallacy lead to frequent deviation from the optimal strategy (always portfolio AABB), only the exposure to risk increases. The expected payout, however, remains unaffected. In fact, the experiment only shows random differences concerning the payouts. Neither consequent herding nor a permanent status-quo bias or a perpetual gambler's fallacy lead to systematically higher payouts than the optimal strategy.

Table 2: Contribution of the Three Treatments to the Objects of Investigation

Treatment	Rational Behavior Hypothesis 1	Herding Hypotheses 2, 3, 4	Status-Quo Bias Hypotheses 5, 6	Gambler's Fallacy Hypotheses 7, 8
Treatment 1 VIEW+0%OPT	X	X	X	X
Treatment 2 NOVIEW+0%OPT	X	X		X
Treatment 3 VIEW+100%OPT	X		X	X

2.3 Herding

As are investigating herding, or the influence of the guru effect, the subjects must be given the opportunity to copy the portfolio decisions of the majority or those of the most successful fellow investor in each period. This results in an experiment that is structured in multiple periods. The portfolios can be rearranged at no cost before the start of each period.

In Treatment 1 (VIEW+0%OPT), the portfolios of each subject as well as their investment success are published in a ranking. Thus, before they decide whether or not to restructure their own portfolio, the subjects are given insight into their fellow investors' portfolio choices in the past period and into the portfolio of the most successful subject. This allows the subjects to follow the majority or the most successful investor (guru). In Treatment 2 (NOVIEW+0%OPT), the subjects do not receive any information about the other subjects' behavior or their investment success. They are informed solely of their own success and therefore do not have the option of following a guru or the majority because neither are detectable (see Figure 3 and Table 2).

Given the numerous empirical findings on the occurrence of herding in the financial market, we expect the portfolios to assimilate during Treatment 1 (VIEW+0%OPT). Thus our second hypothesis reads as follows.

Hypothesis 2: The subjects are not going to converge in the 15 periods of Treatment 1 (VIEW+0%OPT) and will not form a herd.

Since the investment behavior and success of the other subjects cannot be observed in Treatment 2, we expect the subjects to be less distracted from the rational strategy (always portfolio AABB). Thus our third hypothesis reads as follows.

Hypothesis 3: The average deviation from the rational strategy (always portfolio AABB) is not going to be stronger in Treatment 1 (VIEW+0%OPT) than in Treatment 2 (NOVIEW+0%OPT).

If deviations from the rational strategy occur more often and are stronger in Treatment 1 (VIEW+0%OPT), this should show in the average exposure to risk.

Hypothesis 4: The average exposure to risk is not significantly higher in Treatment 1 (VIEW+0%OPT) than in Treatment 2 (NOVIEW+0%OPT).

Figure 3: Interrelation of the Three Treatments
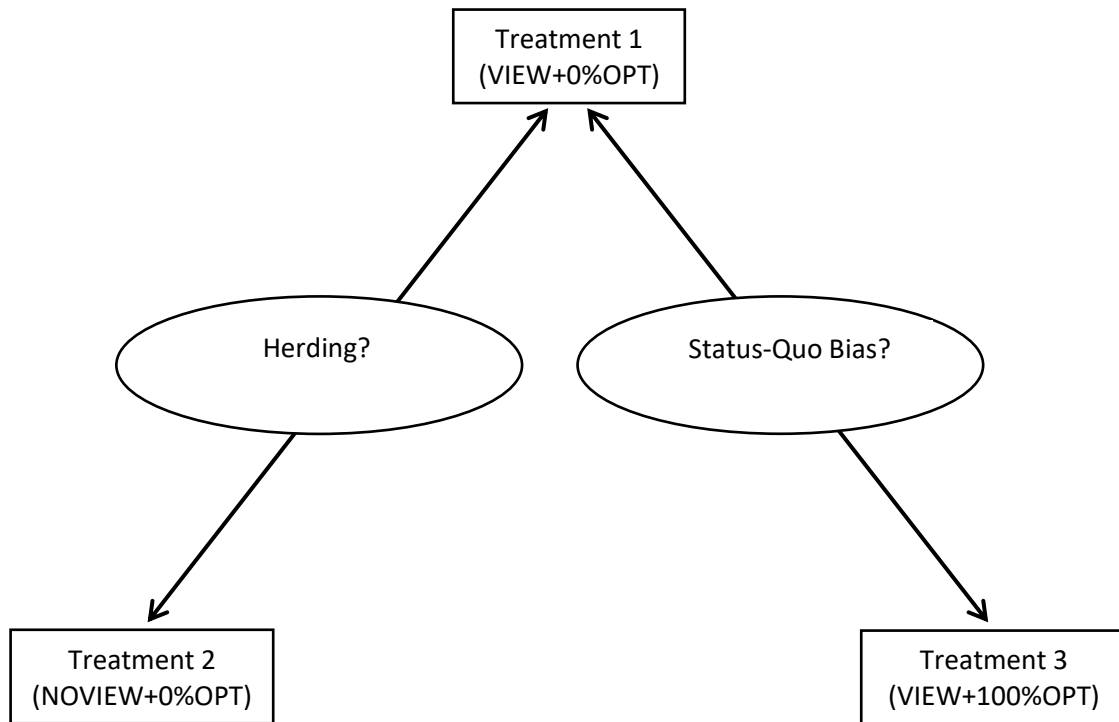
2.4 Status-Quo Bias

To reveal, or not, the presence of status-quo bias, we equip the subjects with different stocks in Treatment 1 (VIEW+0%OPT): 25% of the subjects each start with portfolio AAAA, portfolio AAAB, portfolio ABBB, and portfolio BBBB; 0% of the subjects start with the optimal portfolio (AABB). The subjects were allowed to reassemble their portfolios before the first period. In Treatment 3 (VIEW+100%OPT), 100% of the subjects receive the optimal portfolio (AABB) at the beginning of the experiment, which they can again reassemble before it starts (see Figure 3 and Table 2). As the status-quo bias has often been empirically established, we assume that the optimal portfolio (AABB) is more frequently selected during the 15 periods of Treatment 3 (VIEW+100%OPT) than during Treatment 1 (VIEW+0%OPT).

Hypothesis 5: The average deviation from the rational strategy (always portfolio AABB) will not be stronger in Treatment 1 (VIEW+0%OPT) than in Treatment 3 (VIEW+100%OPT).

If the deviations from the rational strategy are stronger and occur more often in Treatment 1 (VIEW+0%OPT), this should show in the average exposure to risk.

Hypothesis 6: The average exposure to risk will not be higher in Treatment 1 (VIEW+0%OPT) than in Treatment 3 (VIEW+100%OPT).

2.5 The Gambler's Fallacy

To detect whether the subjects have fallen prey to the gambler's fallacy, we ask them in all three treatments, as well as between Periods 4 and 5 and Periods 10 and 11, about the reasons for their portfolio choices (in Period 5 and Period 11). The tendency toward the gambler's fallacy can take effect in all three treatments and is therefore investigated in all three (see Table 2).

As the gambler's fallacy is an often observed phenomenon, we expect the subjects to try to detect patterns in the history of random events, which do not exist. For example: "After Stock A has generated a high income return, I will put my faith in Stock B" or "After Stock B has lastly returned no income, I will choose Stock B." Responses like this are evidence of the gambler's fallacy. We expect this type of answer to be given often.

Hypothesis 7: The gambler's fallacy is not going to be one of the main reasons for certain portfolio choices.

Evidence of the gambler's fallacy can also be discovered by looking at the history of the game. If a positive (negative) event for Stock A (Stock B) frequently leads to a reduced (increased) interest in Stock A (Stock B) in the following period, the influence of gambler's fallacy can be inferred.

Hypothesis 8: A positive (negative) income return in the current period does not reduce (increases) the popularity of this stock in the next period.

2.6 Capture of Risk Attitude and Conduction of the Experiment

The exclusive rational strategy for risk-averse investors is to always choose portfolio AABB. Therefore, deviations from the rational strategy will be identified only if risk-averse subjects are admitted to the experiment. To this end, we tested each subject according to Holt and Laury (2002) and cleared the starting field of risk-neutral and risk-loving subjects. To ensure that the task was fully understood by all subjects, we asked them control questions. Only those who answered all control questions correctly were admitted to the experiment. The complete instructions and control questions can be found in Appendix 1.

The experiment was conducted from 19 May 2016 to 27 May 2016 at the Ostfalia Laboratory for Experimental Economic Research (Ostfalia Labor für experimentelle Wirtschaftsforschung *OLEW*) of the Ostfalia University of Applied Sciences in Wolfsburg. One-hundred-eighty-eight subjects took part in the experiment, 38 of whom showed to be risk-neutral or risk-loving. One-hundred-fifty subjects showed risk-averse behavior and thus were admitted to the actual experiment. Fifty-three subjects were exposed to Treatment 1, 46 subjects to Treatment 2, and 51 took part in Treatment 3. The subjects are students of the Ostfalia University of Applied Sciences in Wolfsburg, 84 of whom study at the Faculty of Business (44.7%), 28 at the Faculty for Health Services (14.9%), and 76 at the Faculty of Automotive Engineering (40.4%). Sixteen sessions were conducted in total. Ten to twelve subjects took part in each session.

The experiment was implemented in z-Tree (see Fischbacher, 2007). At the Ostfalia Lab, we used 12 workspaces, each equipped with a monitor, with a wall separating the subjects. The experiments were consistently overseen by a game master to prevent the subjects from communicating with each other or using unauthorized devices (like smartphones). The subjects did not receive a general show-up fee. When assessing their willingness to take risks, an average of €2.18 was paid out to each subject. The actual experiment resulted in a payout of €21.89 on average. In total, the subjects received an average payout of €24.07. The highest payout was €31.85, the lowest was €17.40. The experiment lasted 45 minutes on average. The payout can therefore be deemed highly attractive. All subjects appeared to concentrate and seemed motivated.

3 Results

3.1 Rational Strategy

The results of the experiments partly meet the expectations but also reveal some surprising facts. Hypothesis 1 states that the subjects are going to behave rationally, which means that they are going to exclusively choose the portfolio with the minimum variance (AABB). Tables 3 and 4 clearly show that Hypothesis 1 must be rejected. The optimal portfolio (AABB) is the most frequently chosen alternative in all three treatments, but more than 60% of all portfolio choices can be viewed as less than completely optimal (Table 3). This is also evident when analyzing the exposure to risk (average standard deviation of the portfolios). The t-test (one-sample mean-comparison) shows that, in all three treatments, portfolios with standard deviations significantly higher than the standard deviation of the optimal portfolio are chosen (Table 4). Many subjects thus exhibit non-rational investment behavior. On the other hand, however, extreme portfolios (AAAA or BBBB) were chosen in only 21.4% of all portfolio decisions.

Table 3: Percentage Distribution of the Portfolios in the Three Treatments

Treatment	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios
Rational Strategy for all Three Treatments	0%	0%	100%	0%	0%
Treatment 1 (VIEW+0%OPT)	11.68%	18.73%	39.75%	19.36%	10.44%
Treatment 2 (NOVIEW+0%OPT)	8.68%	22.45%	31.89%	22.46%	14.49%
Treatment 3 (VIEW+100%OPT)	8.36%	21.69%	40.93%	18.29%	10.71%
Total	9.63%	20.88%	37.74%	19.95%	11.77%

Table 4: Exposure to Risk (Average Standard Deviation of the Portfolios)

Treatment	Rational Strategy: Average SD	Actual: Average SD	P-Value
Treatment 1 (VIEW+0%OPT)	9.9	11.37	0.0000***
Treatment 2 (NOVIEW+0%OPT)	9.9	11.49	0.0000***
Treatment 3 (VIEW+100%OPT)	9.9	11.29	0.0000***
Total	9.9	11.40	0.0000***

*** = significant with an error rate of 1%; ** = significant with an error rate of 5%; * = significant with an error rate of 10%; SD = standard deviation.

3.2 Herding

Herding should be reflected in either a quick or gradual assimilation of the subjects' decisions. We therefore examine whether the decisions made by the subjects converge. Table 5 shows that the subjects' behavior continues to be fragmented until the last period, that is, no herding occurs.[10]

Table 5: Percentage Distribution of the Portfolios at the End of the Game in Treatment 1 (VIEW+0%OPT)

Session	Percentage of AAAA Portfolios in Period 15	Percentage of AAAB Portfolios in Period 15	Percentage of AABB Portfolios in Period 15	Percentage of ABBB Portfolios in Period 15	Percentage of BBBB Portfolios in Period 15
Session 1	30.0%	10.0%	50.0%	10.0%	0.0%
Session 2	33.3%	33.3%	22.2%	11.1%	0.0%
Session 3	11.1%	33.3%	44.4%	0.0%	11.1%
Session 4	30.0%	20.0%	40.0%	0.0%	10.0%
Session 5	0.0%	42.9%	28.6%	14.3%	14.3%
Session 6	37.5%	25.0%	0.0%	12.5%	25.0%

In the next step, we compare decisions for the portfolio that was preferred by most subjects to the portfolio that was chosen against the majority opinion. We conclude that the subjects chose the portfolio based on majority opinion in the previous period for a total of 320 times. The portfolio that the majority did not prefer in the previous period was chosen 422 times. Since herding can also be a temporary phenomenon, it makes sense to show the frequencies separately according to periods and to check the significance of the frequencies using a Chi-square goodness-of-fit test (Table 6). In Period 2 we can see, for instance, that 30 subjects follow the majority opinion of the previous period in choosing their portfolio, while only 23 subjects do not do so. However, this difference is insignificant. In those periods with significant

---

[10] The course of the six sessions of Treatment 1 are set out in Tables A-1 to A-6 in Appendix 2.

deviations (Periods 8, 9, 10, 11, and 13), those decisions not following the majority opinion always outbalance those that do. Hence, we cannot establish significant herding, evidenced by following the majority opinion, at any point during the experiment.

Table 6: Portfolio Decisions According to and Opposing the Majority Opinion in Treatment 1

Period	Number of Portfolios According to the Majority Opinion in Previous Period	Number of Portfolios Opposing the Majority Opinion in Previous Period	P-Value
2	30	23	0.336
3	25	28	0.680
4	31	22	0.216
5	24	29	0.492
6	24	29	0.492
7	27	26	0.891
8	20	33	0.074*
9	18	35	0.020**
10	18	35	0.020**
11	19	34	0.039**
12	22	31	0.216
13	15	38	0.002***
14	22	31	0.216
15	25	28	0.680
total	320	422	

*** = significant with an error rate of 1%; ** = significant with an error rate of 5%; * = significant with an error rate of 10%.

Neither is there much evidence of a "guru effect" (see Table 7). The guru obviously only impresses the subjects slightly. In 212 cases, their decision follows the previous leader's portfolio; they do not do so in 530 cases. Only in Period 7 are there more subjects following the guru's opinion than those not doing so. The Chi-square goodness-of-fit test shows, however, that the difference is insignificant. In all periods – except for Periods 3 and 7 – the decisions not to choose the portfolio of the most successful subject is observed significantly more often.

Overall, we conclude that there is no herding in Treatment 1 (VIEW+0%OPT). Neither the majority opinion nor the opinion of the leading subject (guru) has much impression on the subjects. Hypothesis 2 states that the subjects are not going to converge in the 15 periods of Treatment 1 and will not form a herd. Hypothesis 2 cannot be rejected.

Contrary to our expectations, the portfolio with the minimum variance is chosen significantly more often in Treatment 1 (VIEW+0%OPT) than in Treatment 2 (NOVIEW+0%OPT). We assumed that, by observing and following their fellow subjects' behavior, the subjects would frequently deviate from choosing the optimal stock combination AABB. In Treatment 2, where the other subjects' behavior cannot be observed, herding is generally impossible. As a matter of fact, the contrary turns out to be the case. While the portfolio with minimum variance (AABB) was chosen in 39.8% of all cases in Treatment 1 (VIEW+0%OPT), it was selected in only 31.9% of all cases in Treatment 2 (NOVIEW+0%OPT). The Wilcoxon-Mann-Whitney test shows

that this unexpected difference is even statistically significant. The p-value is 0.021. Hypothesis 3 states that the average deviation from the rational strategy (always portfolio AABB) is not going to be stronger in Treatment 1 than in Treatment 2. Hypothesis 3 cannot be rejected.

Table 7: Portfolio Decisions Following and Not Following the Guru in Treatment 1

Period	Number of Portfolios Corresponding to the Leading Subject of the Previous Period	Number of Portfolios Opposing the Leading Subject of the Previous Period	P-Value
2	6	47	0.000***
3	22	31	0.216
4	12	41	0.000***
5	14	39	0.001***
6	19	34	0.039**
7	28	25	0.680
8	17	36	0.009***
9	7	46	0.000***
10	19	34	0.039**
11	12	41	0.000***
12	9	44	0.000***
13	11	42	0.000***
14	16	37	0.004***
15	20	33	0.074*
total	212	530	

*** = significant with an error rate of 1%; ** = significant with an error rate of 5%; * = significant with an error rate of 10%.

We had expected that the possibility of following other subjects would lead to a significantly higher exposure to risk (average standard deviation of the portfolios) in Treatment 1 (VIEW+0%OPT) than in Treatment 2 (NOVIEW+0%OPT). Since herding did not develop, risk exposure did not increase. The average standard deviation of the portfolios in Treatment 1 (VIEW+0%OPT) was 11.37. The average standard deviation of the portfolios in Treatment 2 (NOVIEW+0%OPT) was 11.49. According to the Wilcoxon-Mann-Whitney test, this difference is not relevant. The p-value is 0.5485. Hypothesis 4 states that the average exposure to risk is not significantly higher in Treatment 1 (VIEW+0%OPT) than in Treatment 2 (NOVIEW+0%OPT). Hence, Hypothesis 4 cannot be discarded.

As an interim result, we conclude that the subjects do not behave rationally in making most of their portfolio choices. The optimal portfolio (AABB) is chosen in only 30 - 40% of all cases (39.8% in Treatment 1, 31.9% in Treatment 2). Herding is clearly not responsible for this. There is neither a lasting orientation toward the portfolio structure chosen by a majority of subjects nor a lasting orientation toward the portfolio structure of the most successful subject.

3.3 Status-Quo Bias

To discover, or not, the presence of any status-quo bias, we now compare Treatment 1 with Treatment 3. In Treatment 1 (VIEW+0%OPT), 25% of the subjects are given portfolio AAAA, 25% portfolio AAAB, 25% portfolio ABBB, and 25% portfolio BBBB before the start of the experiment. No subject receives the optimal portfolio AABB (see Table A-7 in Appendix 3). In Treatment 3 (VIEW+100%OPT), every subject received the portfolio with the minimum variance (AABB) (see Table A-9 in Appendix 3). The subjects are permitted to reassemble their portfolios before the beginning of the first period.

We expected a tendency to follow the status-quo, as has often been found in the literature. That is, in Treatment 3 (VIEW+100%OPT), where all subjects start with the optimal portfolio (AABB), the optimal portfolio structure should have been selected more often than in Treatment 1 (VIEW+0%OPT), in which no subject was provided with an ideally structured portfolio. As a matter of fact, the subjects did not retain their initial portfolios. In Treatment 3 (VIEW+100%OPT), 39.2% of the subjects reassembled their portfolio before the start of the first period, resulting in only 60.8% of the portfolios having the ideal structure (AABB) before the start of the first period (see Figure 5 and Table A-9 in Appendix 3). Over the course of the game, this percentage drops even further.

Figure 4: Percental Distribution of the Portfolios in Treatment 1 (VIEW+0%OPT)
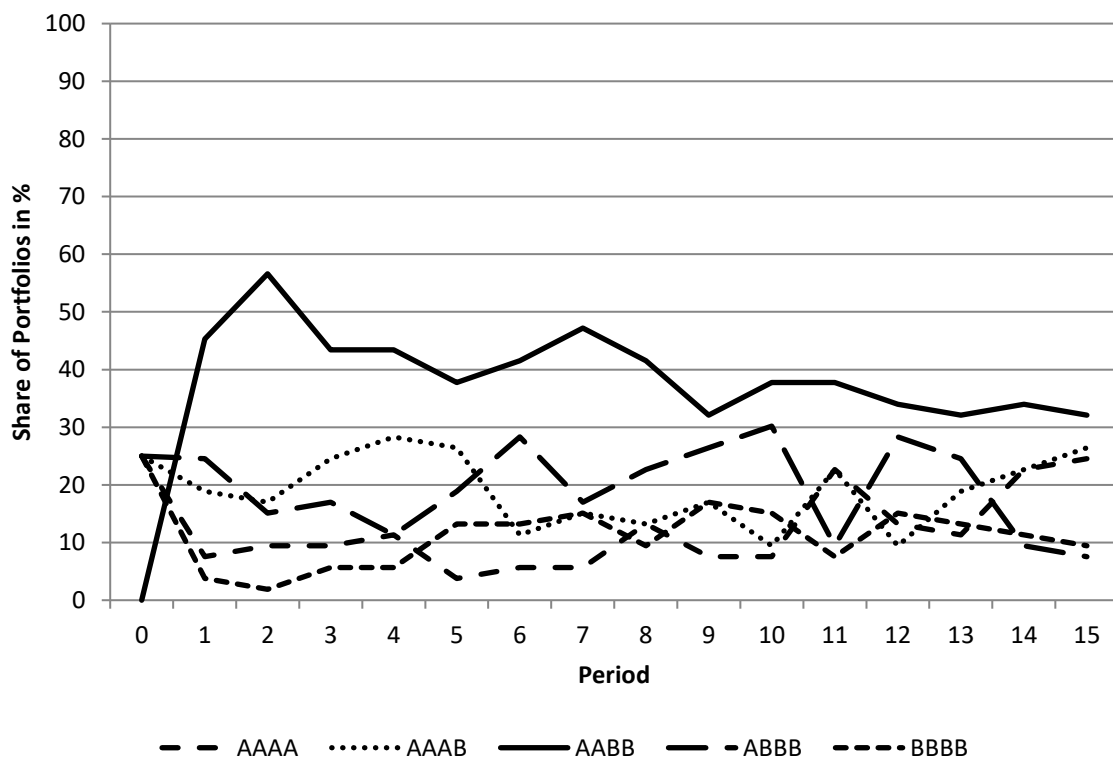
Figure 5: Percental Distribution of the Portfolios in Treatment 3 (VIEW+100%OPT)
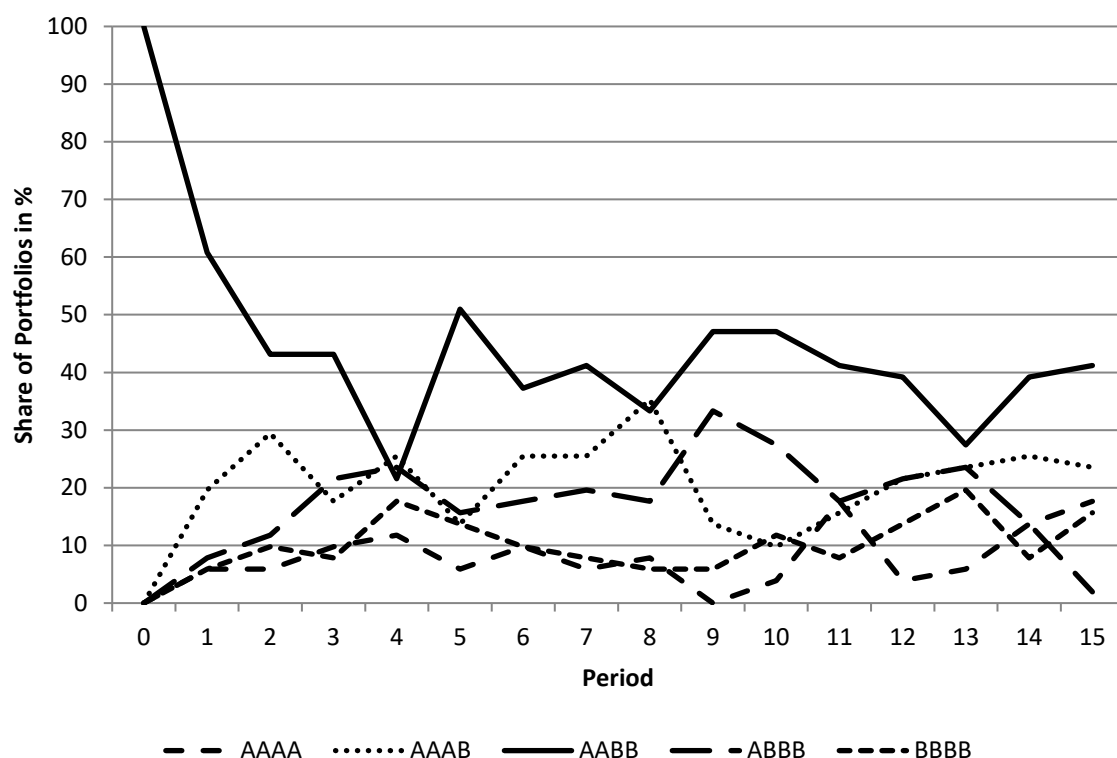
In Treatment 1 (VIEW+0%OPT), the share of the optimal portfolio (AABB) increases to 45.3% even before the first period and ranges between 45% and 30% from Period 3 to Period 15 (see Figure 4 and Table A-7 in Appendix 3).

In Treatment 1 (VIEW+0%OPT), 39.8% of all portfolios have the optimal structure. In Treatment 3 (VIEW+100%OPT), only 40.9% of portfolios with this structure remain. This difference is small and non-relevant according to the Wilcoxon-Mann-Whitney test. The p-value is 0.6626. Hypothesis 5 states that the average deviation from the rational strategy (always portfolio AABB) will not be stronger in Treatment 1 (VIEW+0%OPT) than in Treatment 3 (VIEW+100%OPT). Hypothesis 5 cannot be discarded.

We assumed that, given the status-quo bias in Treatment 3 (VIEW+100%OPT), the optimal portfolio (AABB) would be chosen more often than in Treatment 1 (VIEW+0%OPT), which could have resulted in a significantly lower exposure to risk. As a matter of fact, the average exposures to risk (standard deviation) are 11.37 in Treatment 1 and 11.29 in Treatment 3. According to the Wilcoxon-Mann-Whitney test, this difference is not relevant. The p-value is 0.9741. Hypothesis 6 states that the average exposure to risk will not be higher in Treatment 1 (VIEW+0%OPT) than in Treatment 3 (VIEW+100%OPT). Hypothesis 6 cannot be rejected.

As an intermediate result, we conclude that most subjects do not behave rationally when compiling their portfolios; however, this deviation from the rational strategy is not explained by status-quo bias. Hence, we cannot confirm the result by Brown and Kagel (2009).

3.4 The Gambler's Fallacy

It is plausible to believe that subjects inferred from past random events that a certain sequence of future events would occur, even though this is a nonsensical interpretation of independent random events. To see if this was indeed the case, we assess the reasons the subjects gave for their portfolio choices at the beginning of Period 5 and at the beginning of Period 11. In the experiment, the subjects were asked to provide the following information: "Please give a short explanation for your decision in Period 5 (Period 11)! This explanation does not affect your result! You can openly state your considerations." We expected to receive answers demonstrating that the subjects had indeed fallen prey to the gambler's fallacy.[11]

We differentiate between the following clusters of reasons: (1) Rational consideration (suitable orientation toward the expected value of the income return and the risk); (2) herding (following the majority of investors); (3) herding (following the most successful investor, the guru); (4) status-quo bias (following the present portfolio); (5) the gambler's fallacy (following the history of random events); (6) incomprehensible explanations;[12] and (7) other reasons.[13]

Figure 6: Percentage of the Named Reasons for the Portfolio Choices

---
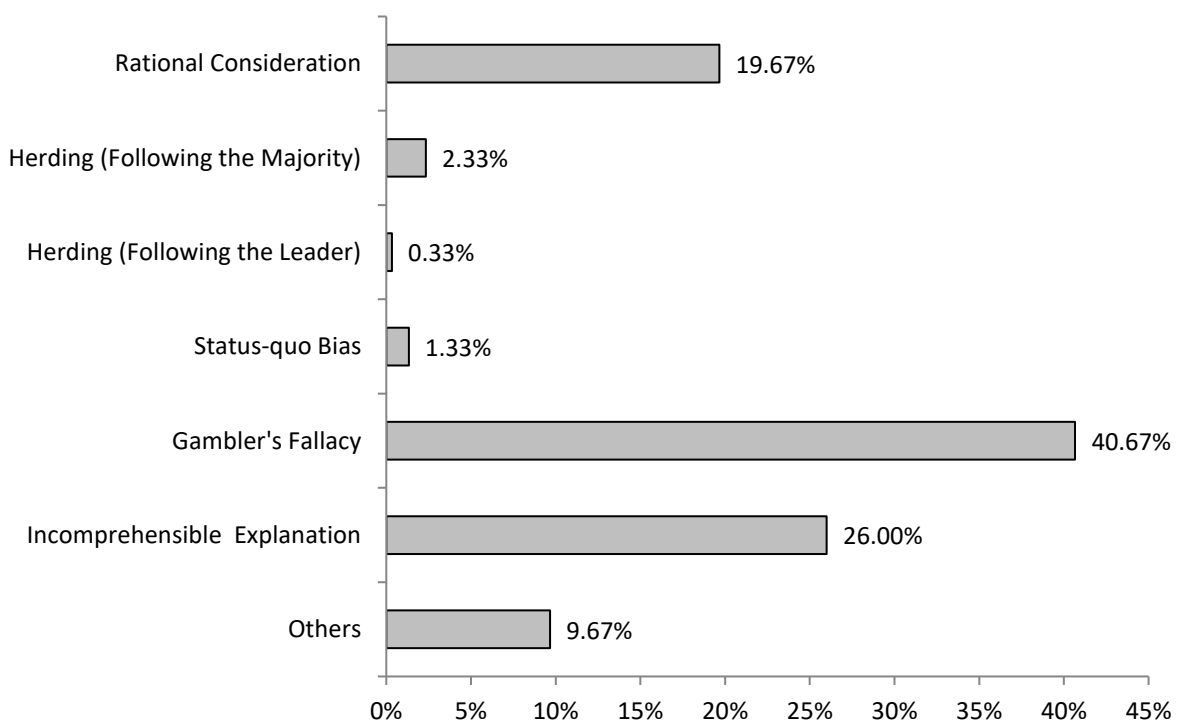
[11] Examples of answers demonstrating the gambler's fallacy: "One of the stocks did not show a positive value development which is why I assume that this stock develops positively in the next period." "The value of Stock A has developed more positively than Stock B which is why I hope that Stock B now develops positively." "The performance of Stock B in period 10 was 0 and Stock A showed an added value of 7, which is why I assume that Stock B could gain in the next period." "With Stock A having developed positively, I now invest in Stock B."
[12] Incomprehensible explanations are, for instance, filling the answer box with random letters or only numbers.
[13] Other reasons are, for example: "I have to take a higher risk to achieve a higher ranking."

It can be concluded that subjects often tried to forecast future random events from past random events. This consideration was detected in 40.67% of the answers (see Figure 6). The frequency of this reasoning significantly exceeds a barely accidental distribution of answers. The Chi-square goodness-of-fit test shows a p-value of 0.000. Hypothesis 7 states that the gambler's fallacy is not going to be one of the main reasons for certain portfolio choices. Hypothesis 7 is rejected. It appears that the gambler's fallacy contributes significantly to irrational portfolio choices.

The question remains whether a behavioral pattern explaining the gambler's fallacy can be directly explained by analyzing the portfolio choices. The sequence of random events (income return of Stock A and Stock B) was structured by coin toss in advance and then taken as a basis in all treatments (see Table 8).

Table 8: Sequence of Random Events and Resulting Gambler's Fallacy

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Random Event of Stock A	±0	+7	±0	+7	+7	±0	±0	+7	±0	±0	+7	+7	±0	±0	+7
Random Event of Stock B	+7	+7	+7	±0	+7	±0	+7	±0	±0	+7	±0	+7	+7	±0	±0
Preference Gambler's Fallacy		A	A	A	B	B	B	A	B	B	A	B	B	A	A
Actual Overweight of the Stock		A	A	A	B	B	B	A	B	B	A	B	B	A	A
Observations in Field 1 of the 2x2 Matrix		1	2	3				4			5			6	7
Observations in Field 4 of the 2x2 Matrix					1	2	3		4	5		6	7		

The first row of Table 8 lists the periods of the experiment. The second and the third rows list the sequences of random events of Stock A and Stock B (income return of ±0 or +7). The fourth line shows which stock is preferred by the subjects who fall prey to the gambler's fallacy. In Period 1, for instance, Stock A has an income return of ±0 ECU and Stock B has an income return of +7 ECU. This is the development that the subjects can see before Period 2. This results in a preference for Stock A. Before the start of Period 3, the "errant gambler" again prefers Stock A. This is because, in the previous period, both stocks had a positive income return (+7 ECU) but for Stock A, it is the first in succession, while it is the second in succession for Stock B. Before Period 4, the events of Period 2 are repeated and result in yet another preference for Stock A. Period by period, considerations such as these lead to preferences that are listed in the fourth line of the table.

If all portfolios of all subjects contain more Stocks A than Stocks B, we call this an overweight of Stock A. If all portfolios of all subjects contain more Stocks B than Stocks A, we call this an overweight of Stock B. We can establish that Stock A (Stock B) is being overly weighted if the gambler's fallacy leads to a preference for Stock A (Stock B). And this is exactly what occurs in the fifth row of Table 8 (for more specific detail, see Table A-10 in Appendix 4). In seven periods (Periods 2, 3, 4, 8, 11, 14, and 15), we would expect Stock A to be overly weighted according to the gambler's fallacy, which is indeed the case (see sixth row in Table 8). In seven periods (Periods 5, 6, 7, 9, 10, 12, and 13), we would expect Stock B to be overly weighted according to gambler's fallacy, which is also the case (see seventh row in Table 8). In a 2 x 2 matrix on the predictive accuracy of the gambler's fallacy regarding the overweight of stocks A and B in Periods 2–15, the Chi-square goodness-of-fit test shows that a predictive accuracy of 14:0, which is highly significant (p-value = 0.000).

Hypothesis 8 states that a positive (negative) income return in the current period does not reduce (increases) the popularity of this stock in the next period. Hypothesis 8, therefore, must be rejected. It is obvious that the gambler's fallacy affects the subjects' portfolio choices and thereby contributes to the fact that the rational choice (to always chose the portfolio with the minimum variance) is observed in only 37.7% of all decisions made by the subjects.

This result specifically confirms the study by Stöckl *et al.* (2015) that also demonstrates that investment decisions can be influenced by the gambler's fallacy. We can also conclude that the gambler's fallacy is rather unaffected by general conditions. It occurs regardless of whether information is or is not provided about the other subjects' behavior, and also manifests regardless of whether or not subjects start with an optimal portfolio. Stöckl *et al.* (2015), too, affirm the stability of the gambler's fallacy under various modes of communication or in the event of group decisions.

Barron and Leider (2010) also find indications of the gambler's fallacy. However, in their study, the gambler's fallacy does not prove stable if long historic time series of random events are shown to the subjects. Chen, Moskowitz, and Shue (2016) also find indications of gambler's fallacy, which in their study is stronger in inexperienced subjects than in experienced subjects. Furthermore, Suetens, Galbo-Joergensen, and Tyran (2016) show that the gambler's fallacy can change into a hot-hand fallacy if a certain event reoccurs in unusually long streaks. Our results are in accordance with the studies by Huber, Kirchler, and Stöckl (2010) and Ayton and Fischer (2004). These studies show that the gambler's fallacy manifests in the prognosis of random events. But these studies also show that, when estimating people's success, the hot-hand fallacy is predominant.

4 Conclusion

The present experimental study examines subjects' diversification behavior. It focuses on the research question of whether herding (being guided by most fellow investors or by the most successful investor (guru)), status-quo bias, and / or the gambler's fallacy can explain why many subjects maintain clearly underdiversified portfolios. Although much empirical evidence

has been found to explain the influence of these phenomena (herding, status-quo bias, and the gambler's fallacy) on many economic decisions, they have not been experimentally examined in the context of diversification decisions.

This experiment follows the approach by Gubaydullina and Spiwoks (2015): There are only two alternatives for investment (Stock A and Stock B), which can produce only two results. Either they bring an income return of ±0 ECU or a return of +7 ECU per period. Both results occur with a probability of 50%. Stock A's return and Stock B's return are independent events. In this environment, the efficient frontier is reduced to one point (equal mixture of Stock A and Stock B), so that the subjects' degree of risk aversion does not influence optimal portfolio choice.

In Treatment 1 (VIEW+0%OPT), the subjects receive information on their fellow investors in each period. They learn who chose which portfolio and how successful everyone was with their decisions. In Treatment 2 (NOVIEW+0%OPT) this information is not provided. A comparison of Treatment 1 and Treatment 2 is intended to indicate the extent to which subjects are distracted from making optimal portfolio choices due to herding behavior (being guided by most fellow investors or by the most successful investor (guru)). Herding was not observed: The subjects neither followed most of their fellow investors nor did they emulate the most successful investor (guru). We draw the conclusion that herding does not play a significant role in explaining non-optimal portfolio choices.

In Treatment 1 (VIEW+0%OPT), the subjects start with different portfolios. Each of the portfolios AAAA, AAAB, ABBB, and BBBB is given to 25% of the subjects at the start of the experiment. In Treatment 3 (VIEW+100%OPT), 100% of the subjects receive the optimal portfolio (AABB). If status-quo bias is playing a role in diversification decisions, the optimal portfolio should be chosen more often in Treatment 3 (VIEW+100%OPT) and the average exposure to risk should be less than in Treatment 1 (VIEW+0%OPT). However, the experiment did not reveal a significant difference between Treatment 1 and Treatment 3 regarding the choice of optimal portfolios or the average exposure to risk. Hence, status-quo bias does not explain non-optimal portfolio choices. In this respect, our results clearly contradict the study by Brown and Kagel (2009).

Over the course of the experiment, the subjects were asked twice about their reasons for making a certain decision. We conclude from their answers that many subjects are inclined to infer future developments from past random events. If a positive event (+7 ECU) has occurred for Stock A (Stock B), the subjects tend to assume that Stock B (Stock A) will be the optimal choice in the next period. Conversely, if a negative event (±0 ECU) has occurred for Stock A (Stock B), the subjects tend to assume that Stock A (Stock B) will be the optimal choice in the next period. This phenomenon, known as the gambler's fallacy, is clearly visible when analyzing the subjects' portfolio choices. This result particularly confirms the study by Stöckl *et al.* (2015) that shows that investment decisions can be influenced by the gambler's fallacy and that the latter is stable under different general conditions.

Most portfolio choices (62.3%) in this experiment were non-optimal. However, neither herding nor status-quo bias contributed to this irrational behavior. The gambler's fallacy, however, had a substantial influence on the subjects' portfolio choices.

Acknowledgments

References

Agnew, J. / Balduzzi, P. / Sundén, A. (2003): Portfolio Choice and Trading in a Large 401(k) Plan, American Economic Review, 93(1), 193-215.

Anderson, A. (2013): Trading and Under-Diversification, Review of Finance, 17(5), 1-43.

Anderson, C. J. (2003): The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion, Psychological Bulletin, 129(1), 139-167.

Ayton, P. / Fischer, I. (2004): The Hot-Hand Fallacy and the Gambler's Fallacy: Two Faces of Subjective Randomness? Memory and Cognition, 32(8), 1369-1378.

Baltussen, G. / Post, G. T. (2011): Irrational Diversification: An Examination of Individual Portfolio Choice, Journal of Financial Quantitative Analysis, 46(5), 1463-1491.

Banerjee, A. V. (1992): A Simple Model of Herd Behavior, Quarterly Journal of Economics, 107(3), 797-817.

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

Barron, G. / Leider, S. (2010): The Role of Experience in the Gambler's Fallacy, Journal of Behavioral Decision Making, 23(1), 117-129.

Belhoula, M. / Naoui, K. (2011): Herding and Positive Feedback Trading in American Stock Market: A Two-Directional Behavior of Investors, International Journal of Business Management, 6(9), 244-252.

Benartzi, S. (2001): Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock, Journal of Finance, 56(5), 1747-1764.

Benartzi, S. / Thaler, R. H. (2001): Naïve Diversification Strategies in Defined Contribution Saving Plans, American Economic Review, 91(1), 79-98.

Bikhchandani, S. / Hirshleifer, D. / Welch, I. (1998): Learning from the Behavior of Others: Conformity, Fads and Informational Cascades, Journal of Economic Perspectives, 12(3), 151-170.

Blume, M. E. / Friend, I. (1975): The Asset Structure of Individual Portfolios and Some Implications for Utility Functions, Journal of Finance, 30(2), 585-603.

Bode, M. / van Echelpoel, A. / Sievi, C. R. (1994): Multinationale Diversifikation: Viel zitiert, kaum befolgt, Die Bank, 94(4), 202-206.

Boyson, N. M. (2010): Implicit Incentives and Reputational Herding by Hedge Fund Managers, Journal of Empirical Finance, 17(3), 283-299.

Brown, A. L. / Kagel, J. H. (2009): Behavior in a Simplified Stock Market: The Status-Quo Bias, the Disposition Effect and the Ostrich Effect, Annals of Finance, 5(1), 1-14.

Bryant, L. L. / Evans, J. / Bishara, J. (2012): Status Quo Bias in the Mutual Fund Market: Consumer Inertia Amidst Benefitless Fee Increases, Banking and Finance Review, 4(1), 89-107.

Chang, C. H. (2013): The Relationship between the Disposition Effect and Herding Behavior: Evidence from Taiwan's Information Technology Stocks, International Journal of Economics and Management Engineering, 7(7), 1971-1975.

Chang, E. C. / Cheng, J. W. / Khorana, A. (2000): An Examination of Herd Behavior in Equity Markets: An International Perspective, Journal of Banking and Finance, 24(10), 1651-1679.

Charness, G. / Gneezy, U. / Imas, A. (2013): Experimental Methods: Eliciting Risk Preferences, Jorunal of Economic Behavior and Organization, 87, 43-51.

Chen, D. / Moskowitz, T. J. / Shue, K. (2016): Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers and Baseball Umpires, Quarterly Journal of Economics, 131(3), 1181-1242.

Chen, Y. / Wang, C. / Lin, F. (2008): Do Qualified Foreign Institutional Investors Herd in Taiwan's Securities Market? Emerging Markets Finance and Trade, 44(4), 62-74.

Chiang, T. C. / Zheng, D. (2010): An Empirical Analysis of Herd Behavior in Global Stock Markets, Journal of Banking and Finance, 34(8), 1911-1921.

Choe, H. / Kho, B. C. / Stulz, R. M. (1999): Do Foreign Investors Destabilize Stock Markets? The Korean Experience in 1997, Journal of Financial Economics, 54(2), 227-264.

Choi, J. J. / Laibson, D. / Madrian, B. C. / Metrick, A. (2004): For Better or For Worse: Default Effects and 401(k) Savings Behavior, in: Wise, D. A.: Perspectives in the Economics of Aging, University of Chicago Press, 81-121.

Choi, J. J. / Laibson, D. / Madrian, B. C. / Metrick, A. (2009): Reinforcement Learning and Savings Behavior, Journal of Finance, 64(6), 2515-2534.

Choi, S. (2016): Herding Among Local Individual Investors: Evidence from Online and Offline Trading, Economics Letters, 144, 4-6.

Christie, W. G. / Huang, R. D. (1995): Following the Pied Piper: Do Individual Returns Herd Around the Market? Financial Analysts Journal, 51(4), 31-37.

Clotfelter, C. T. / Cook, P. J. (1991): The "Gambler's Fallacy" in Lottery Play, NBER Working Paper, 3769.

Crosetto, P. / Filippin, A. (2013): The "Bomb" Risk Elicitation Task, Journal of Risk Uncertainty, 47(1), 31-65.

Devenow, A. / Welch, I. (1996): Rational Herding in Financial Economics, European Economic Review, 40, 603-615.

Dimmock, S. G. / Kouwenberg, R. / Mitchell, O. S. / Peijnenburg, K. (2016): Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence, Journal of Finance and Economics, 119(3), 559-577.

Dohmen, T. / Falk, A. / Huffman, D. / Sunde, U. / Schupp, J. / Wagner, G. G. (2011): Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences, Journal of the European Economic Association, 9(3), 522-550.

Eckel, C. C. / Grossman, P. J. (2002): Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk, Evolution and Human Behavior, 23(4), 281-295.

Eckel, C. C. / Grossman, P. J. (2008): Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices, Journal of Economic Behavior and Organization, 68(1), 1-17.

Ennis, R. M. / Sebastian, M. D. (2003): A Critical Look at the Case for Hedge Funds, Journal of Portfolio Management, 29(4), 103-112.

Eyster, E. / Weizsäcker, G. (2011): Correlation Neglect in Financial Decision Making, DIW Discussion Paper, 1104.

Fellner, G. / Güth, W. / Maciejovsky, B. (2004): Illusion of Expertise in Portfolio Decisions: An Experimental Approach, Journal of Economic Behavior and Organization, 55(3), 355-376.

Fernandes, D. (2013): The 1/N Rule Revisited: Heterogeneity in the Naïve Diversification Bias, International Journal of Marketing Research, 30(3), 310-313.

Fischbacher, U. (2007): Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, Experimental Economics, 10(2), 171-178.

Freiburg, M. / Grichnik, D. (2013): Institutional Reinvestments in Private Equity Funds as a Double-Edged Sword: The Role of the Status-Quo Bias, Journal of Behavioral Finance, 14(2), 134-148.

Galariotis, E. C. / Rong, W. / Spyrou, S. I. (2015): Herding on Fundamental Information: A Comparative Study, Journal of Banking and Finance 50, 589-598.

Geng, S. (2016): Decision Time, Consideration Time and Status-Quo Bias, Economic Inquiry, 54(1), 433-449.

Gneezy, U. / Potters, J. (1997): An Experiment on Risk Taking and Evaluation Periods, Quarterly Journal of Economics, 112(2), 631-645.

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

Gubaydullina, Z. / Hein, O. / Spiwoks, M. (2011): The Status-Quo Bias of Bond Market Analysts, Journal of Applied Finance and Banking, 1(1), 31-51.

Gubaydullina, Z. / Spiwoks, M. (2015): Correlation Neglect, Naïve Diversification, and Irrelevant Information as Stumbling Blocks for Optimal Diversification, Journal of Finance and Investment Analysis, 4(2), 1-19.

Guiso, L., / Haliassos, M. / Japelli, T. (2002): Household Portfolios, MIT Press, Cambridge.

Hedesstrom, T. M. / Svedsater, H. / Garling, T. (2006): Covariation Neglect among Novice Investors, Journal of Experimental Psychology-App., 12(3), 155-165.

Hein, O. / Schwind, M. / Spiwoks, M. (2008): Frankfurt Artificial Stock Market – A Microscopic Stock Market Model with Heterogeneous Interacting Agents in Small-World Communication Networks, Journal of Economic Interaction and Coordination, 3(1), 59-71.

Hein, O. / Schwind, M. / Spiwoks, M. (2012): Network Centrality and Stock Market Volatility: The Impact of Communication Topologies on Prices, Journal of Finance and Investment Analysis, 1(1), 199-232.

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

Holt, C. A. / Laury, S. K. (2002): Risk Aversion and Incentive Effects, American Economic Review, 92(5), 1644-1655.

Huang, T. C. / Wu, C. C. / Lin, B. H. (2016): Institutional Herding and Risk-Return Relationship, Journal of Business Research, 69(6), 2073-2080.

Huber, J. / Kirchler, M. / Stöckl, T. (2010): The Hot Hand Belief and the Gambler's Fallacy in Investment Decisions under Risk, Theory and Decision, 68(4), 445-462.

Huberman, G. / Sengmueller, P. (2004): Performance and Employer Stock in 401(k) Plans, Review of Finance, 8(3), 403-443.

Inman, J. J. / Zeelenberg, M. (2002): Regret in Repeat Purchase Versus Switching Decisions: The Attenuating Role of Decision Justifiabilit, Journal of Consumer Research, 29(1), 116-128.

Kahneman, D. / Tversky, A. (1982): The Psychology of Preferences, Scientific American, 246(1), 160-173.

Kallir, I. / Sonsino, D. (2009): The Neglect of Correlation in Allocation Decisions, Southern Economic Journal, 75(4), 1045-1066.

Kempf, A. / Ruenzi, S. (2006): Status Quo Bias and the Number of Alternatives: An Empirical Illustration from the Mutual Fund Industry, Journal of Behavioral Finance, 7(4), 204-213.

Keynes, J. M. (1936): The General Theory of Employment, Interest and Money, Macmillan, London.

Kim, W. / Jegadeesh, N. (2010): Do Analysts Herd? An Analysis of Recommendations and Market Reactions, Review of Financial Studies, 23(2), 901-937.

Klemkosky, R. C. (1977): The Impact and Efficiency of Institutional Net Trading Imbalances, Journal of Finance, 32(1), 79-86.

Kraus, A. / Stoll, H. R. (1972): Parallel Trading by Institutional Investors, Journal of Financial Quantitative Analysis, 7(5), 2107-2138.

Kremer, S. / Nautz, D. (2013): Causes and Consequences of Short-Term Institutional Herding, Journal of Banking and Finance, 37(5), 1676-1686.

Lakonishok, J. / Shleifer, A. / Vishny, R. W. (1992): The Impact of Institutional Trading on Stock Prices, Journal of Finance and Economics, 32(1), 23-43.

Lease, R. C. / Lewellen, W. G. / Schlarbaum, G. G. (1974): The Individual Investor: Attributes and Attitudes, Journal of Finance, 29(2), 413-433.

Lejuez, C. W. / Aklina, W. M. / Zvolenskyb, M. J. / Pedullac, C. M. (2003): Evaluation of the Balloon Analogue Risk Task (BART) as a Predictor of Adolescent Real-World Risk-Taking Behaviours, Journal of Adolescence, 26(4), 475-479.

Lin, W. T. / Tsai, S.-C. / Lung, P.-Y. (2013): Investors' Herd Behavior: Rational or Irrational? Asia-Pacific Journal of Financial Studies, 42(5), 755-776.

Lönnqvist, J.-E. / Verkasalo, M. / Walkowitz, G. / Wichardt, P. C. (2015): Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison, Journal of Economical Behavioral Organization, 119, 254-266.

Mackay, C. (1841): Extraordinary Popular Delusions and the Madness of Crowds, Richard Bentley, London.

Markose, S. / Alentorn, A. / Krause, A. (2004): Dynamic Learning, Herding and Guru Effects in Networks, Working Paper, Economics Department, University of Essex.

Markowitz, H. (1952): Portfolio Selection, Journal of Finance, 7(1), 77-91.

Meulbroek, L. (2005): Company Stock in Pension Plans: How Costly Is It? Journal of Law and Economics, 48(2), 443-474.

Morrin, M. / Inman, J. J. / Broniarczyk, S. M. / Nenkov, G. Y. / Reuter, J. (2012): Investing for Retirement: The Moderating Effect of Fund Assortment Size on the 1/N Heuristic, Journal of Marketing Research, 49(4), 537-550.

Nofsinger, J. / Sias, R. (1999): Herding and Feedback Trading by Institutional and Individual Investors, Journal of Finance, 54(6), 2263-2295.

Panchenko, V. / Gerasymchuk, S. / Pavlov, O. V. (2013): Asset Price Dynamics with Local Interactions under Heterogeneous Beliefs, Journal of Economic Dynamics and Control, 37(12), 2623-2642.

Patel, J. / Zeckhauser, R. / Hendricks, D. (1991): The Rationality Struggle: Illustrations from Financial Markets, American Economic Review, 81(2), 232-236.

Polkovnichenko, V. (2005): Household Portfolio Diversification: A Case for Rank-Dependent Preferences, Review of Financial Studies, 18(4), 1467-1502.

Powdthavee, N. / Riyanto, Y. E. (2012): Why Do People Pay for Useless Advice? Implications of Gambler's and Hot-Hand Fallacies in False-Expert Setting, IZA Discussion Paper, 6557.

Rieger, M. O. (2012): Why Do Investors Buy Bad Financial Products? Probability Misestimation and Preferences in Financial Investment Decision, Journal of Behavioral Finance, 13(2), 108-118.

Samuelson, W. / Zeckhauser, R. (1998): Status-Quo Bias in Decision Making, Journal of Risk Uncertainty, 1(1), 7-59.

Scharfstein, D. S. / Stein, J. C. (1990): Herd Behavior and Investment, American Economic Review, 80(3), 465-479.

Sias, R. W. (2004): Institutional Herding, Review of Financial Studies, 17(1), 165-206.

Spiwoks, M. (2004): External Triggered Herding bei Rentenmarkt-Analysten, Finanzmarkt und Portfolio Management, 18(1), 58-83.

Spiwoks, M. / Bizer, K. / Hein, O. (2008): Anchoring Near the Lighthouse: Bond Market Analysts' Behavior Co-ordination by External Signal, European Journal of Economics, Finance and Administrative Sciences, 13, 169-191.

Stöckl, T. / Huber, J. / Kirchler, M. / Lindner, F. (2015): Hot Hand and Gambler's Fallacy in Teams: Evidence from Investment Experiments, Journal of Economic Behavior and Organization, 117, 327-339.

Suetens, S. / Galbo-Joergensen, C. / Tyran, J.-R. (2016): Predicting Lotto Numbers: Experiment on the Gambler's Fallacy and the Hot-Hand Fallacy, Journal of the European Economic Association, 14(3), 584-607.

Sumpter, D. J. T. / Zabzina, N. / Nicolis, S. C. (2012): Six Predictions About the Decision Making of Animal and Human Groups, Managerial and Decision Economics, 33(5-6), 295-309.

Tedeschi, G. / Iori, G. / Gallegati, M. (2009): The Role of Communication and Imitation in Limit Order Markets, European Physical Journal B, 71, 489-497.

Tedeschi, G. / Iori, G. / Gallegati, M. (2012): Herding Effects in Order Driven Markets: The Rise and Fall of Gurus, Journal of Economic Behavior and Organization, 81(1), 82-96.

Tversky, A. / Kahneman, D. (1971): Belief in the Law of Small Numbers, Psychological Bulletin, 76(2), 105-110.

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

Voronkova, S. / Bohl, M. T. (2005): Institutional Traders' Behavior in an Emerging Stock Market: Empirical Evidence on Polish Pension Fund Investors, Journal of Business Finance and Accounting, 32(7-8), 1537-1560.

Walter, A. / Weber, F. M. (2006): Herding in the German Mutual Fund Industry, European Financial Management, 12(3), 375-406.

Weber, E. U. / Siebenmorgen, N. / Weber, M. (2005): Communication Asset Risk: How Name Recognition and the Format of Historic Volatility Information Affect Risk Perception and Investment Decisions, Risk Analysis, 25(3), 597-609.

Wermers, R. (1999): Mutual Fund Herding and the Impact on Stock Prices, Journal of Finance, 54, 581-622.

Yen, H. R. / Chuang, S. C. (2008): The Effect of Incidental Affect on Preference for the Status Quo, Journal of the Academy of Marketing Science, 36(4), 522-537.

Zeelenberg, M. / van den Bos, K. / van Dijk, E. / Pieters, R. (2002): The Inaction Effect in the Psychology of Regret, Journal of Personality and Social Psychology, 82(3), 314-327.

Appendix 1: Instructions, Control Questions, Screenshot, Assessment of Risk Attitude

Instructions (Treatment 1 and Treatment 3)

<u>The Game</u>

By investing in stocks, you can benefit from the income return. There are two securities (Stock A and Stock B) to choose from. In each period of the game, the value of Stock A is changed by +0 experimental currency units (ECU) or +7 experimental currency units (ECU); both events occur with a probability of 50%. The same is true for Stock B. Both securities have an expected value of +3.50 ECU per period. The performances of both stocks are independent random events. 15 periods will be played in total and in each period, you have 4 units at your disposal that must be invested. The 4 units can be invested in the following portfolio combinations:

Portfolio 1: AAAA
Portfolio 2: AAAB
Portfolio 3: AABB
Portfolio 4: ABBB
Portfolio 5: BBBB

The game starts at period 0. At period 0, you have the possibility to restructure the portfolio that was randomly assigned to you. The performance of the stocks for the single periods were previously determined by 15 x two coin tosses (one toss for Stock A and one toss for Stock B). "Heads" signifies a good period (+7 ECU) and "tails" signifies a weak period (+0 ECU). You can earn up to 420 ECU in the 15 periods.

For maximum transparency, you will be shown your results and all subjects will be ranked after each period. The ranking is established according to the total earnings. You will therefore be able to compare the performance of your portfolio to the performance of the other subjects' portfolios. You can also earn up to €3.85 in a lottery, detailed information on which you will receive in due course.

<u>The Payout</u>

In the 15 periods, you can earn up to 420 ECU with the securities. 1 ECU equals €0.10. The maximum payout is €42.00 (420 x 0.10). Up to €3.85 from the lottery are added to this sum. In total, you can earn up to €45.85. You will receive your money at the end of the experiment.

<u>Please note:</u>
- Please keep quiet during the experiment!
- Please do not look at your seatmate's monitor!
- Auxiliary devices (calculators, smartphones etc.) are not allowed. All electronic devices must be switched off!

Instructions (Treatment 2)

<u>The Game</u>

By investing in stocks, you can benefit from the income return. There are two securities (Stock A and Stock B) to choose from. In each period of the game, the value of Stock A is changed by +0 experimental currency units (ECU) or +7 experimental currency units (ECU); both events occur with a probability of 50%. The same is true for Stock B. Both securities have an expected value of +3.50 ECU per period. The performances of both stocks are independent random events. 15 periods will be played in total and in each period, you have 4 units at your disposal that must be invested. The 4 units can be invested in the following portfolio combinations:

Portfolio 1: AAAA
Portfolio 2: AAAB
Portfolio 3: AABB
Portfolio 4: ABBB
Portfolio 5: BBBB

The game starts at period 0. At period 0, you have the possibility to restructure the portfolio that was randomly assigned to you. The performance of the stocks for the single periods were previously determined by 15 x two coin tosses (one toss for Stock A and one toss for Stock B). "Heads" signifies a good period (+7 ECU) and "tails" signifies a weak period (+0 ECU). You can earn up to 420 ECU in the 15 periods. For maximum transparency, you will be shown your results after each period. You can also earn up to €3.85 in a lottery, detailed information on which you will receive in due course.

<u>The Payout</u>

In the 15 periods, you can earn up to 420 ECU with the securities. 1 ECU equals €0.10. The maximum payout is €42.00 (420 x 0.10). Up to €3.85 from the lottery are added to this sum. In total, you can earn up to €45.85. You will receive your money at the end of the experiment.

<u>Please Note:</u>
• Please keep quiet during the experiment!
• Please do not look at your seatmate's monitor!
• Auxiliary devices (calculators, smartphones etc.) are not allowed. All electronic devices must be switched off!

Control Questions

Control questions (tick the box):

1.   What is your task in this game?

Ο Solving mathematical problems.

Ο Investing in stocks and taking part in a lottery. (correct)

Ο Giving economic forecasts.

2.   How many different securities are there to choose from and how many free stocks do you receive?

Ο There are 4 different securities to choose from and I receive 2 free stocks.

Ο There are 2 different securities to choose from and I receive 2 free stocks.

Ο There are 2 different securities to choose from and I receive 4 free stocks. (correct)

3.   On what does the payout depend in the 15 periods?

Ο On the dividend payouts.

Ο On the performance of the stocks. (correct)

Ο On the DAX market trend.

4.   How many different combinations of the portfolio are possible?

Ο 2

Ο 4

Ο 5 (correct)

Screenshot Treatment 1 (recreated to increase readability)

You are subject 10. Your result in period 1:

The performance of Stock A in Period 1 is	The performance of Stock B in Period 1 is	Your payout for Period 1 is	Your total payout so far amounts to
±0 ECU	+7 ECU	14 ECU	14 ECU

Previous results of all subjects:

Ranking according to Previous Success	Subjects	Pay-out so far	Last Portfolio Composition
1	3	28 ECU	BBBB
1	5	28 ECU	BBBB
3	4	21 ECU	ABBB
4	7	14 ECU	AABB
4	9	14 ECU	AABB
4	10	14 ECU	AABB
7	1	7 ECU	AAAB
7	6	7 ECU	AAAB
9	2	0 ECU	AAAA
9	8	0 ECU	AAAA

There are still two stocks to choose between (Stock A and Stock B). The value developments for both stocks are independent random processes with the possible results +7 ECU (good period development) and ±0 ECU (unprofitable period development).

You receive 4 stocks. You can choose whether to receive 4 Stocks A, 3 Stocks A and one Stock B, 2 Stocks A and 2 Stocks B, one Stock A and three Stocks B or 4 Stocks B. The value developments of the stocks in period 2 were previously determined by two coin tosses (one toss for Stock A and one toss for Stock B). "Heads" means a good development in the period whereas tails stands for a negative period development (±0 ECU).

Do you want to reallocate your portfolio? If not, please click on "O.K."! If so, please choose your preferred portfolio and then click on "O.K."!

I choose: ☉ 4 Stocks A

☉ 3 Stocks A + 1 Stock B

☉ 2 Stocks A + 2 Stocks B

☉ 1 Stock A + 3 Stocks B

☉ 4 Stocks B

O.K.

Instructions to Determine the Risk Preference

Each decision is a choice between "version A" and "version B". Each version is comparable to a lottery with different payouts and different probabilities of occurrence.

You have 10 decisions to make and enter. One of these decisions will be considered to determine your payout from the lottery as follows: After you entered all your decisions, a ten-sided dice is thrown to select one of the 10 decisions. Each of the decisions therefore has a 10% probability of being chosen. The selected lottery (A or B) is then played. The probability of occurrence is simulated by an urn containing table tennis balls: In an urn with 10 table tennis balls the number of orange balls determines the probability of the higher payout.

Example for decision no. 8: In an urn with 10 table tennis balls are 8 orange-colored and 2 white balls. The probability that a randomly picked ball is orange is therefore 80%. If the picked ball is orange, you will receive €2.00 in version A and €3.85 in version B. If the picked ball is white, you will receive €1.60 in version A and €0.10 in version B. So: You make 10 decisions (either A or B), one of them is randomly selected (with a dice) and played (with an urn and 10 table tennis balls) – the result will determine your payout from the lottery. Please answer the following control questions about the lottery before making any decisions.

| Risk level | Version A: | | | | Version B: | | | | Your Decision A or B |
	p(€2.00)	Payout	p(€1.60)	Payout	p(€3.85)	Payout	p(€0.10)	Payout	
1	10%	€2.00	90%	€1.60	10%	€3.85	90%	€0.10	
2	20%	€2.00	80%	€1.60	20%	€3.85	80%	€0.10	
3	30%	€2.00	70%	€1.60	30%	€3.85	70%	€0.10	
4	40%	€2.00	60%	€1.60	40%	€3.85	60%	€0.10	
5	50%	€2.00	50%	€1.60	50%	€3.85	50%	€0.10	
6	60%	€2.00	40%	€1.60	60%	€3.85	40%	€0.10	
7	70%	€2.00	30%	€1.60	70%	€3.85	30%	€0.10	
8	80%	€2.00	20%	€1.60	80%	€3.85	20%	€0.10	
9	90%	€2.00	10%	€1.60	90%	€3.85	10%	€0.10	
10	100%	€2.00	0%	€1.60	100%	€3.85	0%	€0.10	

Control Questions to Determine the Risk Preference

Control questions (tick the box):

1. What is the minimum and the maximum payout in the lottery?

   Ο The minimum payout is €0.00 and the maximum payout is €1.60.

   Ο The minimum payout is €0.10 and the maximum payout is €3.85. (correct)

   Ο The minimum payout is €0.10 and the maximum payout is €1.60.

2. If the dice selects the 7th decision and you choose version A and have drawn a white table tennis ball from the urn, what is your payout?

   Ο €0.00

   Ο €2.00

   Ο €1.60 (correct)

3. How many white table tennis balls are in the urn if the dice chooses the tenth decision?

   Ο 10

   Ο 0 (correct)

   Ο 5

4. How many orange table tennis balls are in the urn if the dice chooses the fourth decision?

   Ο 6

   Ο 0

   Ο 4 (correct)

Appendix 2: Results of Treatment 1 per Session

Table A-1: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 1)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distrib.	25%	25%	0%	25%	25%	-	-
Period 1	20.0%	0.0%	60.0%	20.0%	0.0%	-	-
Period 2	10.0%	20.0%	70.0%	0.0%	0.0%	AABB	BBBA
Period 3	20.0%	10.0%	30.0%	10.0%	30.0%	AABB	AABB
Period 4	20.0%	10.0%	40.0%	20.0%	10.0%	AABB, BBBB	BBBB
Period 5	0.0%	10.0%	50.0%	10.0%	30.0%	AABB	AAAA
Period 6	0.0%	10.0%	60.0%	20.0%	10.0%	AABB	AAAB
Period 7	0.0%	10.0%	80.0%	0.0%	10.0%	AABB	AABB
Period 8	10.0%	20.0%	40.0%	10.0%	20.0%	AABB	AABB
Period 9	10.0%	0.0%	40.0%	10.0%	40.0%	AABB	BBBA
Period 10	10.0%	10.0%	30.0%	20.0%	30.0%	AABB, BBBB	BBBB, AABB
Period 11	20.0%	10.0%	50.0%	10.0%	10.0%	AABB, BBBB	BBBB
Period 12	20.0%	0.0%	50.0%	20.0%	10.0%	AABB	AAAA
Period 13	20.0%	10.0%	40.0%	20.0%	10.0%	AABB	BBBB
Period 14	30.0%	0.0%	40.0%	30.0%	0.0%	AABB	AAAA
Period 15	30.0%	10.0%	50.0%	10.0%	0.0%	AABB	AAAA

In Session 1 of Treatment 1, 60% of the subjects have already settled on portfolio AABB. This number increases to 70% in Period 2. Subject to some variations, it increases to 80% in Period 7. In Period 10, however, it recedes to 30%.

Table A-2: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 2)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distribution	25%	25%	0%	25%	25%	-	-
Period 1	0.0%	22.2%	44.4%	33.3%	0.0%	-	-
Period 2	0.0%	0.0%	66.7%	33.3%	0.0%	AABB	ABBB
Period 3	0.0%	11.1%	55.6%	33.3%	0.0%	AABB	AABB, ABBB
Period 4	11.1%	11.1%	55.6%	22.2%	0.0%	AABB	ABBB
Period 5	0.0%	11.1%	55.6%	22.2%	11.1%	AABB	ABBB
Period 6	11.1%	0.0%	44.4%	44.4%	0.0%	AABB	AABB, ABBB
Period 7	11.1%	0.0%	55.6%	33.3%	0.0%	AABB, ABBB	AABB, ABBB
Period 8	11.1%	22.2%	44.4%	11.1%	11.1%	AABB	ABBB
Period 9	11.1%	22.2%	33.3%	22.2%	11.1%	AABB	AAAB
Period 10	11.1%	11.1%	44.4%	22.2%	11.1%	AABB	AABB
Period 11	22.2%	22.2%	44.4%	11.1%	0.0%	AABB	BBBB, ABBB
Period 12	0.0%	11.1%	44.4%	33.3%	11.1%	AABB	AAAA
Period 13	11.1%	11.1%	11.1%	55.6%	11.1%	AABB	BBBB
Period 14	22.2%	44.4%	22.2%	0.0%	11.1%	ABBB	BBBB, ABBB
Period 15	33.3%	33.3%	22.2%	11.1%	0.0%	AAAB	BBBB, AAAB

In Session 2 of Treatment 1 we again observe a trend toward the stock combination AABB. In the first period, 44% of the subjects chose this portfolio structure, and 67% do so in the second period. Subsequently, however, these decisions recede to 11% in Period 13, only to end at 22% in the last period.

Table A-3: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 3)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distribution	25%	25%	0%	25%	25%	-	-
Period 1	11.1%	33.3%	33.3%	11.1%	11.1%	-	-
Period 2	11.1%	0.0%	77.8%	11.1%	0.0%	AAAB, AABB	BBBB
Period 3	11.1%	44.4%	44.4%	0.0%	0.0%	AABB	AAAA
Period 4	0.0%	22.2%	66.7%	0.0%	11.1%	AAAB, AABB	AAAA, AAAB, AABB
Period 5	0.0%	11.1%	44.4%	22.2%	22.2%	AABB	AAAB
Period 6	0.0%	0.0%	33.3%	22.2%	44.4%	AABB	BBBB
Period 7	0.0%	22.2%	55.6%	0.0%	22.2%	BBBB	BBBB
Period 8	11.1%	0.0%	55.6%	11.1%	22.2%	AABB	AAAB, AABB, BBBB
Period 9	11.1%	11.1%	22.2%	44.4%	11.1%	AABB	AAAA
Period 10	0.0%	0.0%	33.3%	44.4%	22.2%	ABBB	BBBB
Period 11	33.3%	0.0%	33.3%	22.2%	11.1%	ABBB	BBBB
Period 12	11.1%	11.1%	66.7%	0.0%	11.1%	AABB	AAAA, BBBB
Period 13	0.0%	0.0%	55.6%	22.2%	22.2%	AABB	AAAA, BBBB
Period 14	11.1%	22.2%	55.6%	0.0%	11.1%	ABBB	BBBB
Period 15	11.1%	33.3%	44.4%	0.0%	11.1%	AAAB	AAAA, AAAB

In Session 3 of Treatment 1, an increasing number of subjects choose portfolio BBBB. None of the subjects in Period 3 decides for this combination. In Period 4, 11% make this decision, and as much as 22% do so in Period 5, and even 44% in Period 6. This development, however, is not continued but collapses rather quickly.

Table A-4: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 4)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distribution	25%	25%	0%	25%	25%	-	-
Period 1	0.0%	30.0%	40.0%	30.0%	0.0%	-	-
Period 2	10.0%	30.0%	50.0%	10.0%	0.0%	AABB	ABBB
Period 3	0.0%	20.0%	50.0%	30.0%	0.0%	AABB	AAAA, AABB, ABBB
Period 4	10.0%	40.0%	40.0%	10.0%	0.0%	AABB	ABBB
Period 5	0.0%	40.0%	30.0%	20.0%	10.0%	AAAB, AABB	AAAB
Period 6	0.0%	30.0%	40.0%	20.0%	10.0%	AAAB	AAAB
Period 7	10.0%	50.0%	20.0%	10.0%	10.0%	AABB	AAAB
Period 8	10.0%	20.0%	55.0%	20.0%	0.0%	AAAB	BBBB
Period 9	0.0%	30.0%	30.0%	30.0%	10.0%	AABB	AAAA
Period 10	0.0%	10.0%	40.0%	40.0%	10.0%	AAAB, AABB, ABBB	AABB
Period 11	10.0%	40.0%	40.0%	10.0%	0.0%	AABB, ABBB	AABB
Period 12	10.0%	10.0%	20.0%	40.0%	20.0%	AAAB, AABB	AAAA
Period 13	0.0%	50.0%	40.0%	0.0%	10.0%	ABBB	BBBB
Period 14	30.0%	20.0%	40.0%	0.0%	10.0%	AAAB	AABB
Period 15	30.0%	20.0%	40.0%	0.0%	10.0%	AABB	AAAA

In Session 4 of Treatment 1, similar results are achieved. The portfolio structure ABBB is increasingly considered by the subjects. In Period 7, this portfolio is selected by 10% of the subjects, and by 20%, 30%, and 40% percent in the subsequent periods. This development stops abruptly after that. In Periods 13 to 15, this combination is no longer chosen.

Table A-5: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 5)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distribution	25%	25%	0%	25%	25%	-	-
Period 1	0.0%	14.3%	71.4%	14.3%	0.0%	-	-
Period 2	0.0%	28.6%	57.1%	14.3%	0.0%	AABB	ABBB
Period 3	0.0%	28.6%	42.9%	28.6%	0.0%	AABB	AAAB
Period 4	0.0%	28.6%	42.9%	14.3%	14.3%	AABB	ABBB
Period 5	0.0%	71.4%	14.3%	14.3%	0.0%	AABB	AAAB
Period 6	0.0%	14.3%	57.1%	28.6%	0.0%	AAAB	AAAB
Period 7	0.0%	0.0%	57.1%	28.6%	14.3%	AABB	ABBB
Period 8	0.0%	14.3%	28.6%	57.1%	0.0%	AABB	ABBB
Period 9	0.0%	14.3%	42.9%	42.9%	0.0%	ABBB	AAAB
Period 10	0.0%	14.3%	28.6%	57.1%	0.0%	AABB, ABBB	AABB
Period 11	14.3%	42.9%	28.6%	0.0%	14.3%	ABBB	AAAB, ABBB
Period 12	14.3%	28.6%	0.0%	57.1%	0.0%	AAAB	AAAB
Period 13	0.0%	14.3%	42.9%	42.9%	0.0%	ABBB	AAAB, ABBB
Period 14	0.0%	14.3%	28.6%	28.6%	28.6%	AABB, ABBB	ABBB
Period 15	0.0%	42.9%	28.6%	14.3%	14.3%	AABB, ABBB, BBBB	AAAB, ABBB

In Session 5 of Treatment 1, the combination ABBB is again popular for some time. The number of subjects deciding for this portfolio increases from Period 5 (14%), over Periods 6 and 7 (29%) to Period 8 with 57%. Subject to some variations, this choice recedes to 14% in Period 15.

Table A-6: Percentage Distribution of the Portfolios in the Game (Treatment 1, Session 6)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios	Most Frequently Chosen Portfolio(s) in the Previous Period	Portfolio(s) of the Leader(s) in the Previous Period
Initial Distribution	25%	25%	0%	25%	25%	-	-
Period 1	12.5%	12.5%	25.0%	37.5%	12.5%	-	-
Period 2	25.0%	25.0%	12.5%	25.0%	12.5%	AABB	BBBB
Period 3	25.0%	37.5%	37.5%	0.0%	0.0%	AAAA, AAAB, ABBB	ABBB
Period 4	25.0%	62.5%	12.5%	0.0%	0.0%	AAAB, AABB	AABB
Period 5	25.0%	25.0%	25.0%	25.0%	0.0%	AAAB	AAAB
Period 6	25.0%	12.5%	12.5%	37.5%	12.5%	AAAA, AAAB, AABB, ABBB	AAAA
Period 7	12.5%	0.0%	12.5%	37.5%	37.5%	ABBB	ABBB
Period 8	37.5%	0.0%	25.0%	37.5%	0.0%	ABBB, BBBB	BBBB
Period 9	12.5%	25.0%	25.0%	12.5%	25.0%	AAAA, ABBB	AABB
Period 10	25.0%	12.5%	50.0%	0.0%	12.5%	AAAB, AABB, BBBB	BBBB
Period 11	37.5%	25.0%	25.0%	0.0%	12.5%	AABB	BBBB
Period 12	25.0%	0.0%	12.5%	25.0%	37.5%	AAAA	AAAA
Period 13	37.5%	25.0%	0.0%	12.5%	25.0%	BBBB	BBBB
Period 14	37.5%	37.5%	12.5%	0.0%	12.5%	AAAA	BBBB
Period 15	37.5%	25.0%	0.0%	12.5%	25.0%	AAAA, AAAB	AAAA

In Session 6 of Treatment 1, no herding can be established for portfolio AAAB. In Period 1, this combination is chosen by 13% of the subjects, by 25% in Period 2, by 38% in Period 3, and by as much as 63% in Period 4. In the following period, this development suddenly stops. As early as Period 7, this portfolio is no longer chosen by any of the subjects.

Appendix 3: Percentage Distribution of the Portfolios in the Game

Table A-7: Percentage Distribution of the Portfolios in the Game (Treatment 1)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios
Initial Distri-bution	25%	25%	0%	25%	25%
Period 1	7.5%	18.9%	45.3%	24.5%	3.8%
Period 2	9.4%	17.0%	56.6%	15.1%	1.9%
Period 3	9.4%	24.5%	43.4%	17.0%	5.7%
Period 4	11.3%	28.3%	43.4%	11.3%	5.7%
Period 5	3.8%	26.4%	37.7%	18.9%	13.2%
Period 6	5.7%	11.3%	41.5%	28.3%	13.2%
Period 7	5.7%	15.1%	47.2%	17.0%	15.1%
Period 8	13.2%	13.2%	41.5%	22.6%	9.4%
Period 9	7.5%	17.0%	32.1%	26.4%	17.0%
Period 10	7.5%	9.4%	37.7%	30.2%	15.1%
Period 11	22.6%	22.6%	37.7%	9.4%	7.5%
Period 12	13.2%	9.4%	34.0%	28.3%	15.1%
Period 13	11.3%	18.9%	32.1%	24.5%	13.2%
Period 14	22.6%	22.6%	34.0%	9.4%	11.3%
Period 15	24.5%	26.4%	32.1%	7.5%	9.4%

Table A-8: Percentage Distribution of the Portfolios in the Game (Treatment 2)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios
Initial Distribution	25%	25%	0%	25%	25%
Period 1	2.2%	15.2%	52.2%	21.7%	8.7%
Period 2	8.7%	30.4%	17.4%	34.8%	8.7%
Period 3	4.3%	28.3%	32.6%	21.7%	13.0%
Period 4	10.9%	34.8%	15.2%	19.6%	19.6%
Period 5	8.7%	13.0%	50.0%	13.0%	15.2%
Period 6	15.2%	13.0%	39.1%	17.4%	15.2%
Period 7	13.0%	19.6%	30.4%	28.3%	8.7%
Period 8	4.3%	39.1%	28.3%	10.9%	17.4%
Period 9	8.7%	21.7%	21.7%	21.7%	26.1%
Period 10	4.3%	13.0%	37.0%	26.1%	19.6%
Period 11	15.2%	26.1%	26.1%	17.4%	15.2%
Period 12	0.0%	10.9%	39.1%	32.6%	17.4%
Period 13	4.3%	19.6%	28.3%	32.6%	15.2%
Period 14	15.2%	30.4%	26.1%	21.7%	6.5%
Period 15	15.2%	21.7%	34.8%	17.4%	10.9%

Table A-9: Percentage Distribution of the Portfolios in the Game (Treatment 3)

Period	Percentage of AAAA Portfolios	Percentage of AAAB Portfolios	Percentage of AABB Portfolios	Percentage of ABBB Portfolios	Percentage of BBBB Portfolios
Initial Distribution	0%	0%	100%	0%	0%
Period 1	5.9%	19.6%	60.8%	7.8%	5.9%
Period 2	5.9%	29.4%	43.1%	11.8%	9.8%
Period 3	9.8%	17.6%	43.1%	21.6%	7.8%
Period 4	11.8%	25.5%	21.6%	23.5%	17.6%
Period 5	5.9%	13.7%	51.0%	15.7%	13.7%
Period 6	9.8%	25.5%	37.3%	17.6%	9.8%
Period 7	5.9%	25.5%	41.2%	19.6%	7.8%
Period 8	7.8%	35.3%	33.3%	17.6%	5.9%
Period 9	0.0%	13.7%	47.1%	33.3%	5.9%
Period 10	3.9%	9.8%	47.1%	27.5%	11.8%
Period 11	17.6%	15.7%	41.2%	17.6%	7.8%
Period 12	3.9%	21.6%	39.2%	21.6%	13.7%
Period 13	5.9%	23.5%	27.5%	23.5%	19.6%
Period 14	13.7%	25.5%	39.2%	13.7%	7.8%
Period 15	17.6%	23.5%	41.2%	2.0%	15.7%

Appendix 4: Further Results

Table A-10: Manifestation of the Gambler's Fallacy in the Portfolio Choices of All Treatments

1	2	3	4	5	6	7	8	9	10	11
Period	Random Event of the Previous Period for Stock A	Random Event of the Previous Period for Stock B	Expected Preference with Gambler's Fallacy	Percentage of AAAA Portfolios (W)	Percentage of AAAB Portfolios (X)	Percentage of AABB Portfolios	Percentage of ABBB Portfolios (Y)	Percentage of BBBB Portfolios (Z)	Surplus (W+X) − (Y+Z)	Answers the Expectation
2	±0	+7	A	8.00%	25.33%	40.00%	20.00%	6.67%	6.67%	Yes
3	+7	+7	A	8.00%	23.33%	40.00%	20.00%	8.67%	2.67%	Yes
4	±0	+7	A	11.33%	29.33%	27.33%	18.00%	14.00%	8.67%	Yes
5	+7	±0	B	6.00%	18.00%	46.00%	16.00%	14.00%	-6.00%	Yes
6	+7	+7	B	10.00%	16.67%	39.33%	21.33%	12.67%	-7.33%	Yes
7	±0	±0	B	8.00%	20.00%	40.00%	21.33%	10.67%	-4.00%	Yes
8	±0	+7	A	8.67%	28.67%	34.67%	17.33%	10.67%	9.33%	Yes
9	+7	±0	B	5.33%	17.33%	34.00%	27.33%	16.00%	-20.67%	Yes
10	±0	±0	B	5.33%	10.67%	40.67%	28.00%	15.33%	-27.33%	Yes
11	±0	+7	A	18.67%	21.33%	35.33%	14.67%	10.00%	15.33%	Yes
12	+7	±0	B	6.00%	14.00%	37.33%	27.33%	15.33%	-22.67%	Yes
13	+7	+7	B	7.33%	20.67%	29.33%	26.67%	16.00%	-14.67%	Yes
14	±0	+7	A	17.33%	26.00%	33.33%	14.67%	8.67%	20.00%	Yes
15	±0	±0	A	19.33%	24.00%	36.00%	8.67%	12.00%	22.67%	Yes

Table A-10 shows that the gambler's fallacy did indeed consistently influence the subjects' portfolio choices. The first column lists those periods of the game in which the gambler's fallacy could take effect. No random events existed before Period 1, which is why the gambler's fallacy could not take effect before the portfolio choice in Period 2. The second and third columns of Table A-10 list the random events of the previous periods. The fourth column shows which stock is preferred by the subjects who fall prey to the gambler's fallacy. In Period 1, for instance, Stock A has an income return of ±0 ECU and Stock B has an income return of +7 ECU. This is the development that the subjects can see before Period 2. This results in a preference for Stock A. Before the start of Period 3, the "errant gambler" again prefers Stock A because, in the previous period, both stocks had a positive income return (+7 ECU) but for Stock A, it is the first in succession, while it is the second in succession for Stock B. Before Period 4, the events of Period 2 are repeated and result in yet another preference for Stock A. Period by period, considerations such as these lead to preferences that are listed in the fourth column of Table A-10.

Columns 5, 6, 7, 8, and 9 of Table A-10 show the percentages of the five possible portfolio compilations (AAAA, AAAB, AABB, ABBB and BBBB) for all three treatments (Treatment 1, Treatment 2, and Treatment 3; for detailed results, see Tables A-7, A-8, and A-9 in Appendix

3). Column 10 of Table A-10 displays an unweighted spread, calculated by subtracting the portfolio decisions that chose Stock B (Columns 8 (Y) and 9 (Z)) from those that prefer Stock A (Columns 5 (W) and 6 (X)). We calculate: (W + X) − (Y + Z). In Period 2 this means: (8.00 + 25.33) − (20.00 + 6.67) = 6.67.

If the subtraction results in a positive value, it can be concluded that the subjects preferred Stock A to Stock B when compiling their portfolios. If the subtraction results in a negative value, it can be concluded that the subjects preferred Stock B to Stock A when assembling their portfolios. Consequently, we can expect that a preference for Stock A results in a positive balance and that a preference for Stock B results in a negative balance. In Column 11 of Table A-10, we analyze which periods match our expectations, and we can see that this occurs without exception. When in Column 4 the Stock A (Stock B) is listed, then the balance in Column 10 is positive (negative).

Considering that the subjects fall prey to the gambler's fallacy and accordingly develop preferences for Stock A or Stock B (Column 4 of Table A-10), then we can find a suitable explanation for the surplus of Stock A (positive indication in Column 10) or, respectively, for the surplus of Stock B (negative indication in Column 10) in all periods.

Table A-11: Variances and Standard Deviations of the Five Portfolios Considering the Actual Events for Stock A and Stock B in Experimental Currency Units (ECU)

Period	Events		Performance of Portfolios				
	Stock A	Stock B	AAAA	AAAB	AABB	ABBB	BBBB
1	±0	+7	±0	+7	+14	+21	+28
2	+7	+7	+28	+28	+28	+28	+28
3	±0	+7	±0	+7	+14	+21	+28
4	+7	±0	+28	+21	+14	+7	±0
5	+7	+7	+28	+28	+28	+28	+28
6	±0	±0	±0	±0	±0	±0	±0
7	±0	+7	±0	+7	+14	+21	+28
8	+7	±0	+28	+21	+14	+7	±0
9	±0	±0	±0	±0	±0	±0	±0
10	±0	+7	±0	+7	+14	+21	+28
11	+7	±0	+28	+21	+14	+7	±0
12	+7	+7	+28	+28	+28	+28	+28
13	±0	+7	±0	+7	+14	+21	+28
14	±0	±0	±0	±0	±0	±0	±0
15	+7	±0	+28	+21	+14	+7	±0
		Variance	209.07	115.27	84.00	115.27	209.07
		Stand. Dev.	14.46	10.74	9.17	10.74	14.46

Chapter 4

MEASUREMENT OF RISK PREFERENCE

with Ibrahim Filiz, Markus Spiwoks and Zulia Gubaydullina
contribution Thomas Nahmer: 35%

Abstract

The procedures previously used to determine risk preference (risk-averse, risk-neutral or risk-loving) exhibit a number of weaknesses. In part, they are so complex and sophisticated that the subjects frequently give spontaneous, ill-considered answers. In this way, their actual risk preference can often not be correctly determined. In addition, in this process there are situations and circumstances in which it is not possible to clearly assign subjects to one of the three categories of risk preference. In addition, with the previous approaches, loss aversion – which has an important influence on risk preference – is not taken into consideration, or only insufficiently. We propose here a new procedure to determine risk preference which is (1) extremely simple and clear, which (2) enables unambiguous differentiation between risk-averse, risk-neutral and risk-loving subjects, and which (3) takes the influence of loss aversion on risk preference into account in an appropriate way.

Keywords

JEL Classification

1 Introduction

Markowitz (1952) shows that for risk-averse subjects it usually makes sense to hold diversified securities portfolios. However, there are many empirical findings which reveal that under-diversified portfolios are very frequently held.[1] Experimental economic research examines this contradiction and finds many reasons why suboptimal decisions are frequently made with regard to diversification.[2]

Meaningful experimental results on diversification behavior can normally only be obtained if clarity about the risk preference of the subjects can be achieved: Because that which is meaningful for a risk-averse subject can be complete nonsense for a risk-loving subject, and vice-versa. In the meantime there are a whole range of procedures available to determine risk preference.[3]

In our view, a good procedure for determining risk preference must above all comply with three criteria:

1. It must be a simple and clear procedure.
2. It must be possible to clearly and unambiguously differentiate between risk-averse, risk-neutral and risk-loving subjects.
3. The influence of loss aversion on risk preference should not be neglected.

We consider these three criteria to be key. (1) The procedure has to be simple and clear so that we can really record the risk preferences of the subjects. In the case of complex and confusing decision-making situations, subjects frequently lose patience and then give spontaneous, ill-considered answers. This can sometimes lead to a blurring of their risk preference rather than it being revealed. (2) In the previous approaches used, there are certain combinations of circumstances in which risk-neutral, risk-averse and risk-loving subjects make – with good reason – the same decisions. In that case, it is not possible to differentiate between the three forms of risk preference. (3) As we will show later on, risk preference is significantly determined by the possibility of suffering losses. Procedures to measure risk preference which do not contain the possibility of losses systematically underestimate the proportion of risk-averse subjects.

Our study is divided up into four sections. First, we evaluate the previous approaches against the background of the three criteria we have postulated. In the following chapter we present our new procedure to measure risk preference. Using an experimental investigation we

---

[1] See, for example, Dimmock et al. (2016), Anderson (2013), Hibbert, Lawrence and Prakash (2012), Goetzmann and Kumar (2008), Meulbroek (2005), Polkovnichenko (2005), Huberman and Sengmueller (2004), Agnew, Balduzzi and Sundén (2003), Guiso, Haliassos and Japelli (2002), Benartzi (2001), Benartzi and Thaler (2001), Barber and Odean (2000), Bode, van Echelpoel and Sievi (1994), Blume and Friend (1975), and Lease, Lewellen and Schlarbaum (1974).

[2] See, for example, Filiz et al. (2018), Gubaydullina and Spiwoks (2015), Fernandes (2013), Morrin et al. (2012), Rieger (2012), Eyster and Weizsäcker (2011), Baltussen and Post (2011), Kallir and Sonsino (2009), Hedesstrom, Svedsater and Garling (2006), Fellner, Güth and Maciejovsky (2004), Choi, Laibson and Madrian (2009), Weber, Siebenmorgen and Weber (2005).

[3] See, for example, Lönnqvist et al. (2015), Charness, Gneezy and Imas (2013), Crosetto and Filippin (2013), Eckel and Grossmann (2002, 2008), Lejuez et al. (2002), Holt and Laury (2002), and Gneezy and Potters (1997).

subsequently show that loss aversion should not be neglected when measuring risk prefer-ence. In the final chapter we summarize the most important results of the investigation.

2 The Previous Approaches and their Weaknesses

In the following section we discuss the approaches of Holt and Laury (2002), Eckel and Gross-man (2008), and Crosetto and Filippin (2013). In addition, we briefly consider the approaches used by Lejuez et al. (2002), Gneezy and Potters (1997), the DOSPERT questionnaire created by Weber, Blais and Betz (2002), and the socio-economic panel (Schupp and Wagner, 2002; Wagner, Burkhauser and Behringer, 1993).

2.1 The Multiple Price List Method of Holt and Laury (2002)

In the multiple price list method of Holt and Laury (2002), subjects are asked to make ten decisions choosing between two lotteries in each case (Table 1). As the first decision, lottery A ($2.00 with a probability of 10% or $1.60 with a probability of 90%) is set against lottery B ($3.85 with a probability of 10% and $0.10 with a probability of 90%). The subject has to decide whether they would play lottery A or lottery B. This is followed by the other nine comparisons between lottery A and lottery B. From the sequence of the ten decisions, conclusions about the risk preferences of the subject are then drawn.

Table 1: The Lottery Alternatives of Holt and Laury (2002)

No.	Lottery A				Lottery B			
	Prob.	Event	Prob.	Event	Prob.	Event	Prob.	Event
1	10%	$2.00	90%	$1.60	10%	$3.85	90%	$0.10
2	20%	$2.00	80%	$1.60	20%	$3.85	80%	$0.10
3	30%	$2.00	70%	$1.60	30%	$3.85	70%	$0.10
4	40%	$2.00	60%	$1.60	40%	$3.85	60%	$0.10
5	50%	$2.00	50%	$1.60	50%	$3.85	50%	$0.10
6	60%	$2.00	40%	$1.60	60%	$3.85	40%	$0.10
7	70%	$2.00	30%	$1.60	70%	$3.85	30%	$0.10
8	80%	$2.00	20%	$1.60	80%	$3.85	20%	$0.10
9	90%	$2.00	10%	$1.60	90%	$3.85	10%	$0.10
10	100%	$2.00	0%	$1.60	100%	$3.85	0%	$0.10

Prob. = probability of occurrence; event = random event.

The main problem of this approach is the complexity of the decision-making situation. Neither the expected returns nor the extent of the risk exposure of the alternatives A and B are clearly recognizable for the subjects. Accordingly, many subjects decide randomly or based on a gut feeling. In this situation, it frequently occurs that ten decisions are made where the decision-making process cannot be clearly interpreted. Jacobson and Petrie (2009) as well as Charnes and Viceisza (2011) show that between 55% and 75% of the decision-making processes cannot

be clearly interpreted. Charness et al. (2018) and Dave et al. (2010) also point out additional uncertainties in the interpretation of results.

The approach used by Holt and Laury (2002) becomes somewhat clearer if one considers the expected returns and the risk (standard deviation) of the ten lottery alternatives (Table 2). In the first lottery alternative, lottery A has an expected return of $1.64 and a standard deviation of 0.12. Lottery B has an expected return of $0.48 and a standard deviation of 1.13. Risk-neutral subjects orientate themselves solely towards expected returns and therefore decide in favor of alternative A. A risk-averse subject will also decide in favor of alternative A, because here the expected returns is higher and at the same time the risk is lower than that of alternative B. But how would a risk-loving subject decide? Expected returns would speak for alternative A, but the risk speaks for alternative B. How a risk-loving subject decides therefore depends on the extent of their appetite for risk. Subjects with a great appetite for risk will choose alternative B because the higher risk more than compensates for the lower expected return. Subjects with a mild appetite for risk will choose alternative A because the higher expected return more than compensates for the lower risk.

Table 2: The Expected Returns and Risk (Standard Deviation) of the Lottery Alternatives used by Holt and Laury (2002) and the Preferences of Risk-Neutral, Risk-Averse and Risk-Loving Subjects

No.	Lottery A		Lottery B		Preference Risk-Neutral	Preference Risk-Averse	Preference Risk-Loving
	E(A)	SD	E(B)	SD			
1	1.64	0.12	0.48	1.13	A	A	A or B
2	1.68	0.16	0.85	1.50	A	A	A or B
3	1.72	0.18	1.23	1.72	A	A	A or B
4	1.76	0.20	1.60	1.84	A	A	A or B
5	1.80	0.20	1.98	1.88	B	A or B	B
6	1.84	0.20	2.35	1.84	B	A or B	B
7	1.88	0.18	2.73	1.72	B	A or B	B
8	1.92	0.16	3.10	1.50	B	A or B	B
9	1.96	0.12	3.48	1.13	B	A or B	B
10	2.00	0.00	3.85	0.00	B	B	B

E(A) = expected returns of lottery A; E(B) = expected returns of lottery B; SD = standard deviation.

In the fifth decision, risk-neutral subjects choose alternative B, because the expected return of $1.98 is higher than that of alternative A ($1.80). Risk-loving subjects also choose alternative B because here both expected return and risk are higher than in alternative A. But how will risk-averse subjects react? The expected return would speak for alternative B, but the risk speaks for alternative A. How the subject decides now depends on the extent of their risk aversion. If they are highly risk-averse, they will choose alternative A because the lower risk offsets the lower expected return. If, however, they are only slightly risk-averse, they will decide in favor of alternative B, because the higher expected return more than compensates for the higher risk.

Now the following question arises: How should subjects be classified who always prefer alternative A in the first four decisions and then prefer alternative B in the last six decisions? These can be either risk-neutral, risk-averse or risk-loving subjects (see Table 2). It cannot therefore be guaranteed that they will be unambiguously assigned to one of the three possible categories of risk preference (risk-averse, risk-neutral or risk-loving).

The approach used by Holt and Laury (2002) therefore does not satisfy any of the three requirements which we formulated at the beginning for reliable determining of risk preference: (1) It is complex and unclear. (2) It does not lead to a clear differentiation between risk-neutral, risk-averse and risk-loving subjects. (3) It does not take the possibility of losses into account.

2.2 The Approach used by Eckel and Grossman (2008)

The approach used by Eckel and Grossman (2008) has the advantage that the decision-making situation is significantly clearer than in the case of Holt and Laury (2002). The subjects decide in favor of one of five possible lotteries. In each lottery there are two possible events which each have a probability of occurrence of 50%. From lottery 1 to lottery 5, the expected values rise, as do the risks (Table 3, Figure 2).

In the loss treatment the participants receive $6 for filling in a questionnaire[4] in the run-up to the lottery. They can lose part of this $6 in lottery 4 and all of it in lottery 5. In order to remunerate all subjects uniformly, the expected values are $6 higher in the no-loss treatment. The approach used by Eckel and Grossman (2008) thus also takes the possibility of losses into consideration.

Table 3: Lottery Alternatives in Eckel and Grossman (2008)

No.	Event	Prob	Return Loss	Return No-Loss	E(r) Loss	E(r) No-Loss	Risk SD
1	A	50%	$10	$16	$10	$16	0
	B	50%	$10	$16			
2	A	50%	$18	$24	$12	$18	6
	B	50%	$6	$12			
3	A	50%	$26	$32	$14	$20	12
	B	50%	$2	$8			
4	A	50%	$34	$40	$16	$22	18
	B	50%	$-2	$4			
5	A	50%	$42	$48	$18	$24	24
	B	50%	$-6	$0			

Event = possible random event; Prob = probability of occurrence; Return Loss = payoff of the coincidental events in the loss treatment; Return No-loss = payoff of the coincidental events in the no-loss treatment; E(r) loss = expected value of the payoff in the loss treatment; E(r) no loss = expected value of the payoff in the no-loss treatment; SD = standard deviation.

---

[4] Zuckerman's sensation-seeking scale. See Zuckerman (1979, 1994).

The approach deployed by Eckel and Grossman (2008) is problematic in that the assignment of the subjects to the three categories of risk preference (risk-averse, risk-neutral and risk-loving) is by no means clear. This becomes apparent when one considers that risk-averse, risk-neutral and risk-loving subjects exhibit fundamentally diverging indifference curves. Risk-averse subjects have rising indifference curves, whereas risk-neutral subjects have absolutely horizontal indifference curves and risk-loving subjects have falling indifference curves (Figure 1).

Figure 1: The Form of the Indifference Curves for Risk-Averse, Risk-Neutral and Risk-Loving Subjects
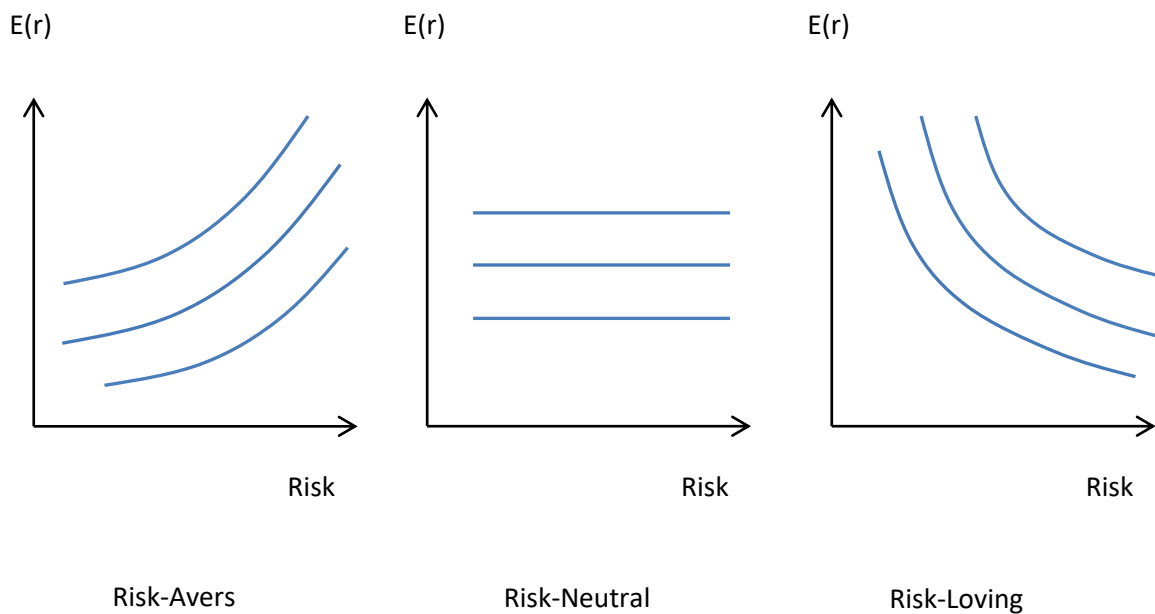
E(r) = expected value of return; risk (standard deviation).

If the space of possibilities which results from the five lotteries is considered, the following becomes recognizable: All of the subjects who choose lottery 5 can be risk-averse as well as risk-neutral or risk-loving (Figure 2).

The approach used by Eckel and Grossman (2008) thus manages to fulfill two of the three criteria we have put forward: It is a simple and clear decision-making situation and the possibility of making losses is also taken into account. However, the unambiguous identification of risk-neutral, risk-averse and risk-loving subjects is not possible.

Figure 2: Space of Possibilities in Eckel and Grossman (2008) as well as the Indifference Curves of a Risk-Averse (Unbroken Grey Line), a Risk-Neutral (Dotted Grey Line) and a Risk-Loving Subject (Dashed Grey Line)
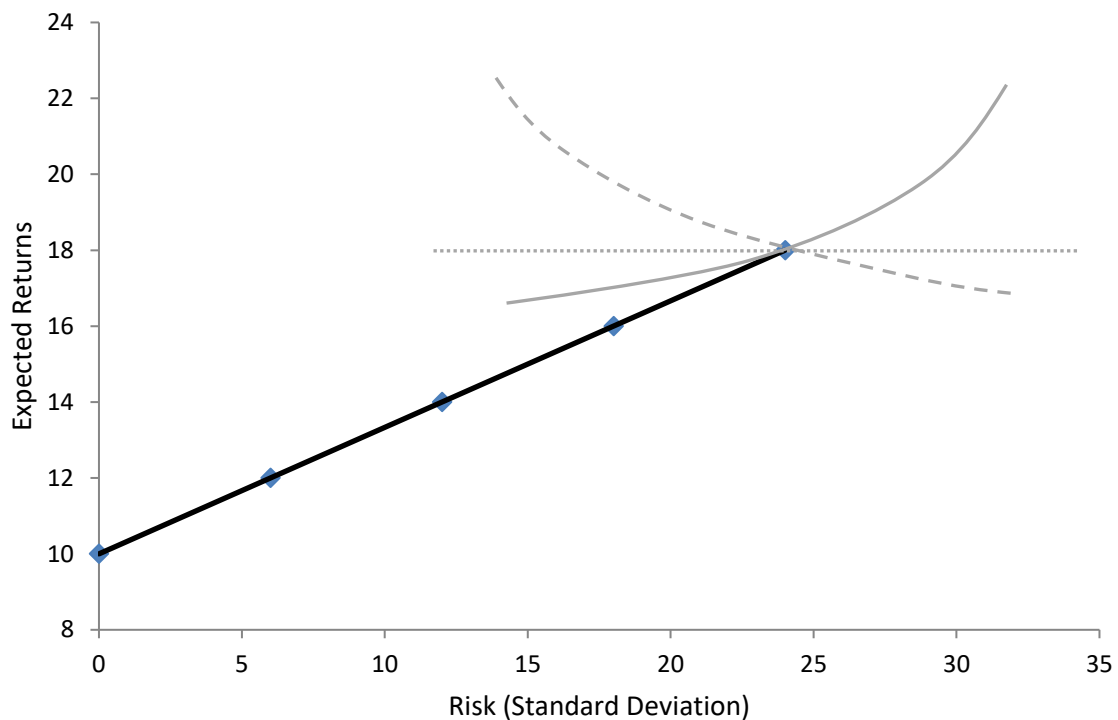
2.3 The Approach used by Crosetto and Filippin (2013)

Crosetto and Filippin (2013) have proposed the most interesting approach yet to determine risk preferences. In this approach, the participants are faced with the following decision-making situation: They have to decide how many of a total of 100 boxes they want to collect. One of the boxes contains a 'bomb'. The participants receive a payoff of €0.10 per box. After they have decided on a number of boxes (static version) or have ended the game by pressing the 'stop button' (dynamic version),[5] a number between 1 and 100 is drawn from an urn. If the number drawn is ≤ the number of collected boxes, the 'bomb' has exploded and the money is gone. If the number drawn is > the number of collected boxes, the subject receives a payment based on the multiplication of the number of boxes collected by €0.10. It can be expected that the subjects want to win as much money as possible. The more boxes they collect, the higher the payoff. At the same time, the risk of encountering the 'bomb' (number drawn ≤ the number of collected boxes) rises. The subjects thus have to weigh up how much risk is meaningful to them. The space of possibilities of this decision-making situation is shown in Figure 3.

---

[5] Crosetto and Filippin deployed a static basic variation and a dynamic variant. In the static variant the subjects only see a picture of 100 boxes and have to decide how many they want to collect. In the dynamic PC version the 100 boxes are shown on the screen. By pressing a start button the participants trigger the collection of one box per second until they press the stop button.

Figure 3: Space of Possibilities in Crosetto and Filippin (2013)
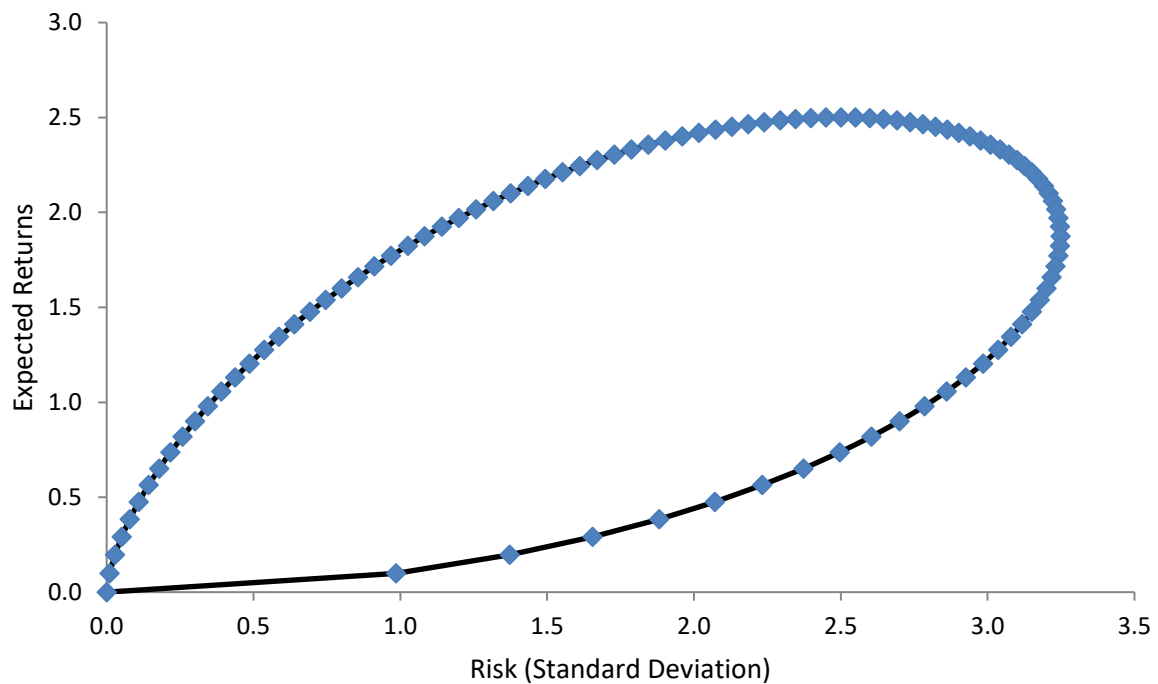
From the first to the 50th box, expected returns rises gradually. At the same time the risk also increases steadily. From the 50th to the 75th boxes, the risk continues to rise, whereas expected returns falls. From the 75th to the 100th boxes, the risk as well as expected returns both decrease. The highest expected return is achieved if one collects exactly 50 boxes. Risk-averse subjects – depending on their risk aversion – will choose between one and 50 boxes. Risk-loving subjects will choose between 50 and 75 boxes. Risk-neutral subjects will always collect exactly 50 boxes, because expected return reaches its maximum level there. The efficient frontier of the space of possibility thus extends from one to 75 boxes. The section from 76 to 100 boxes, however, is the non-efficient part of the space of possibility.

The great advantage of this approach is the enormous clarity of the decision-making situation. In addition, loss opportunities can also be implemented easily, which Crosetto and Filippin (2013) in fact do in one of the treatments.

Nevertheless, some criticism can be made: (1) If a subject collects exactly 50 boxes it is not possible to recognize whether they are risk-averse, risk-neutral or risk-loving. While it is true that all risk-neutral subjects will collect exactly 50 boxes, one cannot conclude that all subjects who collect 50 boxes are risk-neutral. In view of the maximum expected return, slightly risk-averse or slightly risk-loving subjects could also consider 50 boxes to be the most attractive option.[6] (2) The decision-making situation is indeed very clear, but it is not simple. How many subjects recognize that the maximum expected return can be found at exactly 50 boxes? And how many subjects realize what the risk (standard deviation) is for the 100 different

---

[6] Around 14% of the subjects decide to collect exactly 50 boxes. This means that a notable proportion of the subjects cannot be assigned unambiguously to one of the three categories (risk-averse, risk-neutral and risk-loving).

possibilities? A considerable amount of calculating is required to work that out. (3) How should subjects who collect more than 75 boxes be characterized? Those persons who move in the non-efficient part of the space of possibilities are also either risk-averse, risk-neutral or risk-loving. There is no other possibility. However, which of these three alternatives they fit into cannot be said, because each subject who collects more than 75 boxes is obviously not aware of the shape of the space of possibilities.

The three requirements we have put forward for a good process to determine risk preferences are not completely fulfilled here. The decision-making situation is clear, but it is not exactly simple. It is not possible in every case to unambiguously assign subjects to one of the three categories of risk preference (risk-averse, risk-neutral and risk-loving). On the positive side, introducing a risk of loss is simple, which Crosetto and Filippin (2013) in fact do in one of the treatments.

2.4 Further Approaches

The method used by Lejuez et al. (2002) aims to create a relative comparison of risk preference between two or more subjects. However, his aim is not to assign them to one of the three categories of risk preference (risk-averse, risk-neutral and risk-loving). The decision-making situation is designed as follows: A balloon and a pump are shown on a computer screen. With every click of a mouse, the balloon is pumped up a bit more and the participant receives €0.05. Their credit is shown on a temporary account. The subject can stop pumping at any time. If the balloon bursts, the credit accumulated is lost. A total of 90 rounds of the game are played, in which there are three different colored balloons (blue, yellow and orange). The three colors represent different probabilities of bursting. The subjects are only informed that the three different-colored balloons have a different bursting point, and that the balloon can even burst on the first pump. The average number of pumps made is used as an indicator for risk prefer-ence. As no advance information is provided about expected returns and risk, this method is not suitable for assigning subjects to one of the three categories of risk preference: Only a relative comparison between subjects can take place. For example, it can be established that subject A acts more cautiously than subject B. However, whether subject A is risk-averse and subject B is risk-loving remains unclear. Subject A could be strongly risk-averse and subject B could be slightly risk-averse. Or subject A is slightly risk-loving and subject B is highly risk-lov-ing. This remains unclear.

The method used by Gneezy and Potters (1997) examines which proportion of their portfolio subjects invest in a risky asset. To do so, they are asked which proportion of 200 cents they want to bet on in a lottery which there is a probability of two thirds that they will lose the amount and a probability of one third that they will win two and a half times the amount. So if they win they retain the amount they wager plus two and a half times the amount as win-nings. The lottery thus has a positive expected value. A total of nine rounds are played. In treatment H, the participants decide separately for each round which proportion of the 200 cents they want to bet. In treatment L, decisions are made in advance for each of three rounds

of the game. The amount which is wagered thus remains constant for three rounds. Depending on the treatment, the participants are informed about the (aggregated) results after one or three lotteries and then they bet again. It is shown that the average amount placed as a bet in treatment L (decision in advance) is greater than in treatment H (separate decision for each round). The results reveal that an investment period spread over several periods leads to a larger proportion of the investor's assets being invested in a risky asset. In an adapted form, Charness and Gneezy (2010) established that the participants of the experiment would pay in order to have more frequent opportunities to change the composition of their portfolio. However, the structure of the experiment is not suited to assigning the subjects to one of the three categories of risk preference (risk-averse, risk-neutral and risk-loving). Once again, the approach can only be used to establish that subject A acts more cautiously than subject B. The same issue arises as in the case of Lejuez et al. (2002).

Another way of determining individual risk preference is to interview the subjects. A good example of this is the domain-specific risk taking questionnaire (DOSPERT) developed by Weber, Blais and Betz (2002). The questionnaire relates to a large number of high-risk activities or behaviors from five fields: (1) Sports and leisure, (2) health, (3) social issues, (4) ethics, and (5) finances. The questionnaire records the probability of the respondents taking risks, their perception of these risks and of the benefit which might result from the risks taken. A total of 40 topics are evenly distributed over five fields, whereby only the field of finance is subdivided into (a) gambling and (b) investment risks. The participants estimate their own risk preference on a scale from 1 (low-risk) to 5 (high-risk). Assignation to one of the three categories of risk preference (risk-averse, risk-neutral and risk-loving) is not possible on the basis of this questionnaire. Once again, this approach can only be used to establish that subject A acts more cautiously than subject B. The same issues arise as in the case of Lejuez et al. (2002).

Another example of surveying risk preference within the framework of a questionnaire is the socio-economic panel (SOEP). Schupp and Wagner (2002) as well as Wagner, Burkhauser and Behringer (1993) describe the approach used in the questionnaire. The idea is that the interviewees provide information about their general risk preferences. Assignation to one of the three categories of risk preference (risk-averse, risk-neutral and risk-loving) is not possible on the basis of this questionnaire. The same applies to the differentiated versions of the SOEP approach (Schupp and Wagner, 2002; Wagner, Burkhauser and Behringer, 1993). These approaches can only be used to establish that subject A acts more cautiously than subject B. The same issues arise as in the case of Lejuez et al. (2002).

Lönnqvist et al. (2015) examine the time stability of various procedures for the measurement of risk preferences, while Charness, Gneezy and Imas (2013) compare different procedures for the measurement of risk preferences. However, they do not provide a different approach for the identification of risk-neutral, risk-averse and risk-loving subjects.

3 The New Approach

We propose a procedure for differentiating between risk-averse, risk-neutral and risk-loving subjects which is very clear and simple and which makes it possible to assign subjects unambiguously to the three categories of risk preference.

It deals with a decision to choose between two lotteries.[7] The subjects take a card – they can choose between taking a card from pile A or one from pile B. Both piles consist of four playing cards each. In pile A there are two cards which lead to a profit of €+4, and two cards which lead to a profit of €+6 (Figure 4). In pile B there are two cards which lead to no profit €±0, and two cards which lead to a profit of €+10 (Figure 5).

Figure 4: The Four Cards in Pile A

Figure 5: The Four Cards in Pile B

The subjects are informed that the expected return in both piles is identical at €+5. In addition, the subjects are made aware of the fact that pile A leads to results which fluctuate slightly around the expected value (low-risk), while pile B leads to results which fluctuate considerably around the expected value (high-risk). The two piles of cards containing four cards each are not only shown on the screen, but can also be seen as real playing cards on the table of the game leader. The subjects are informed that the pile of cards which they decide for (A or B) will be shuffled and that they then have to take a card. The entire survey is programmed in z-Tree (Fischbacher, 2007). However, we have decided not to program random events in z-Tree, but to carry them out analogously. In this way we want to counteract the possible suspicion that it could be a manipulated random event. The subjects see the playing cards and can be

---

[7] We were inspired by Bechara et al. (1994) here.

sure that there is a probability of exactly 50% that the favorable event (€+6 in pile A and €+10 in pile B) will occur. In addition, they can also be sure that there is a probability of exactly 50% that the unfavorable event (€+4 in pile A and €±0 in pile B) will occur (Table 4).

Table 4: Lottery Alternatives in the New Approach

Pile	Prob	Return	E(r)	Risk (SD)
A	50%	€+4	€+5	1.0
	50%	€+6		
B	50%	€±0	€+5	5.0
	50%	€+10		

Prob = probability; E(r) = expected value of the return; SD = standard deviation.

The shuffling of the cards is left to a machine in order to avoid the suspicion that the game leader has an influence on random events. After the cards have been shuffled, the subjects have to take one of the four cards from the pile they have chosen. They then receive the payment which is noted on this card. Test questions are used to ensure that the subjects understand the circumstances. The instructions for the game, the test questions and selected screenshots can be viewed in the appendix.

The two lotteries lead to a clear space of possibilities (Figure 6) which also permits an unambiguous assignment of the subjects to the three categories of risk preference (risk-averse, risk-neutral and risk-loving).

The space of opportunities consists of only two points. The left point shows the expected return and risk profile of pile A (low-risk). The right point shows the expected return and risk profile of pile B (high-risk).

The decision options for the subjects are:

- I would like to take a card from pile A.
- I would like to take a card from pile B.
- I would like to take a card. I don't mind which pile I take one from.

If one takes into account the form of the indifference curves for risk-averse, risk-neutral and risk-loving subjects (Figure 1), the decision made by the subjects leads to an unambiguous assignment to one of the three categories of risk preference: Risk-averse subjects prefer pile A. Risk-loving subjects prefer pile B. Risk-neutral subjects are indifferent as to whether they choose pile A or B (Figure 6).

Figure 6: Space of Possibilities of the New Approach as well as the Indifference Curves of a Risk-Averse (Unbroken Grey Line), a Risk-Neutral (Dotted Grey Line) and a Risk-Loving Subject (Dashed Grey Line)
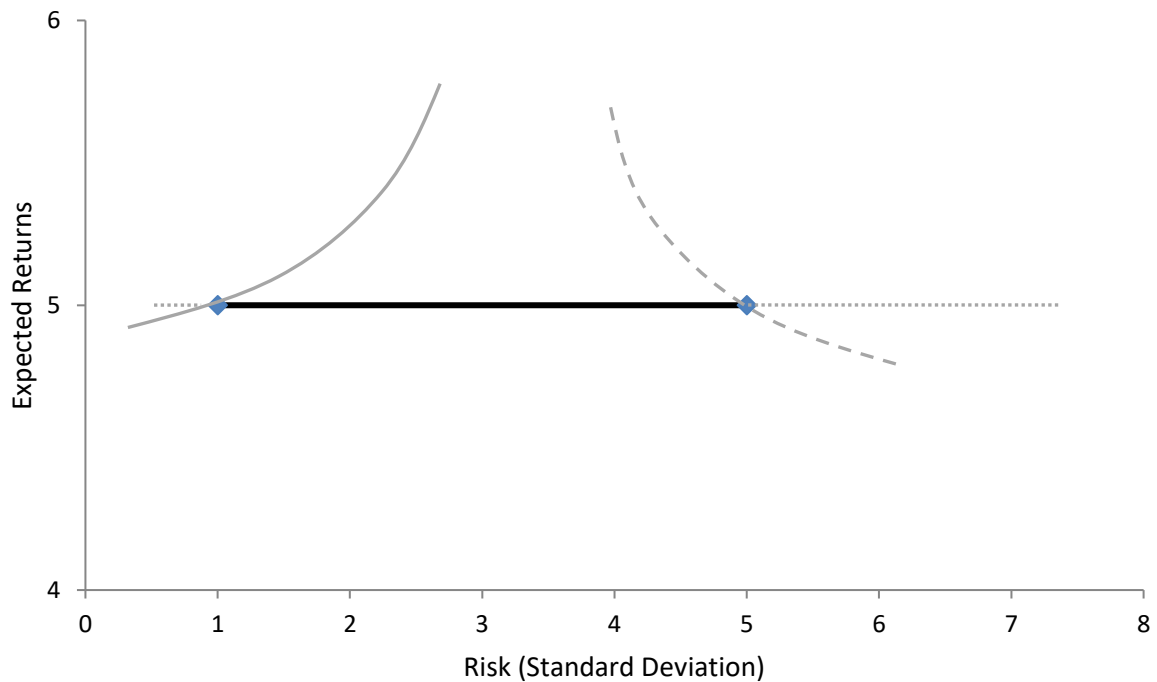
In this way, two out of the three criteria for a suitable procedure to record the risk preference of the test persons are fulfilled: (1) The decision-making situation is very clear and simple. The subjects know precisely which consequences their decision will have. They do not have to decide on the basis of a gut feeling, but can make well thought-out, conscious decisions corresponding to their preferences. (2) The three alternatives (pile A, pile B or indifference as to whether the card is from A or B) permit unambiguous conclusions about the three categories of risk preference (risk-averse, risk-neutral and risk-loving). In the following chapter we will also take the influence of loss aversion on risk preference into account.

4 Taking Loss Aversion into Account in the New Approach

There is hardly another phenomenon in behavioral economics which has been the subject of as much research as loss aversion (for a comprehensive overview see, for example, Kahneman 2011, Chapter 29; see also Rabin 2000; Fehr and Goette 2007; Tom et al., 2007). Frequently, subjects are strongly influenced in their actions by the effort to avoid losses. One would expect that risk preference would also be influenced by the possibility of the threat of losses. However, this presumption has not yet been confirmed. Eckel and Grossman (2008) and Crosetto and Filippin (2013) have both included treatments with the possibility of losses. In spite of this, notable effects on the risk preferences of the subjects could not be observed in either study.

Mukherjee et al. (2017) showed that in the case of small amounts up to $4, a profit had a greater positive influence on the well-being of the participants than an equally high loss had in negative terms. Here, an evaluation scale ranging from 0 (= no effect) to 5 (= very strong effect) was used. In the case of an amount of $25, however, the negative perception of a loss was more intense than the positive feeling of an equally large profit. These results indicate that loss aversion might only have an influence on risk preference when larger amounts are involved. In Eckel and Grossman (2008), losses between $-2 and $-6 can occur. And in Crosetto and Filippin (2013), losses of €-2.50 can occur.

We will now carry out the new approach to establish risk preference in three variations in order to investigate the influence of loss aversion on risk preference in more detail. In Treatment 1, there is no possibility of making a loss. In Treatment 2, a loss of €-2.50 can be made. In Treatment 3, a loss of €-25 can be made (Table 5).

Table 5: Random Events, Expected Values and Standard Deviations in Treatments 1-3

Treatment	Pile	Prob.	Return	E(r)	SD
1	A	50% 50%	€+4 €+6	€+5	1.0
	B	50% 50%	€±0 €+10	€+5	5.0
2	A	50% 50%	€+4 €+6	€+5	1.0
	B	50% 50%	-2.5 €+12.5	€+5	7.5
3	A	50% 50%	€+4 €+6	€+5	1.0
	B	50% 50%	€-25 €+35	€+5	30

Prob. = probability; E(r) = expected value of the return; SD = standard deviation.

The results obtained by Mukherjee et al. (2017), Eckel and Grossman (2008) and Crosetto and Filippin (2013) lead us to expect that there will be no significant differences between Treatment 1 and Treatment 2. The possible losses of €-2.50 are presumably too small to have an influence on the risk preferences of the subjects. The first hypothesis is therefore as follows:

Hypothesis 1: In Treatment 2, not more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 1.

The first null hypothesis which will have to be examined is therefore:

Null hypothesis 1: In Treatment 2, significantly more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 1.

The results of Mukherjee et al. (2017), however, give reason to presume that the danger of losses of €-25 can have an influence on the risk preferences of the subjects. The second hypothesis is therefore as follows:

Hypothesis 2: In Treatment 3, more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 1.

The second null hypothesis to be examined is therefore:

Null hypothesis 2: In Treatment 3, not more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 1.

If the presumption is correct that the possibility of a small loss does not really impress subjects, whereas that of a larger loss has a significant influence on risk preferences, it must also be possible to establish a difference between Treatment 2 and Treatment 3. Our third hypothesis is therefore as follows:

Hypothesis 3: In Treatment 3, more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 2.

The third null hypothesis which will have to be examined is therefore:

Null hypothesis 3: In Treatment 3, not more (less) subjects will prove to be risk-averse (risk-loving) than in Treatment 2.

In our experiment we conduct a between-subjects comparison. A total of 157 students of the Ostfalia University of Applied Sciences in Wolfsburg took part in the experiment. 53 subjects played Treatment 1, 52 subjects played Treatment 2, and 52 subjects played Treatment 3. 53 women (33.76%) and 104 men (66.24%) took part. 72 of the subjects study business management (45.86%), 69 subjects study vehicle construction (43.95%) and 16 students study health care (10.19%). The experiment was carried out from 4-10 April 2018 in the Ostfalia Laboratory for Experimental Economic Research (OLEW) in Wolfsburg in Germany. The experiment is programmed in z-Tree. Only the playout of random events is carried out in an analogue way by taking a card from the respective selected pile.[8]

The actual experiment is preceded by a real effort task. We give the subjects a task which is not enjoyable and which requires a considerable amount of time. The subjects are supposed to view the task as work which is paid for with an appropriate amount (€25). The subjects have to encode a total of 175 three-letter words in sequences of numbers. When they have encoded a word correctly, the next word appears. This real effort task is based on Erkal, Gangadharan and Nikiforakis (2011). In order to make it more demanding, Benndorf, Rau and Solch (2014) change the assignment of numbers to letters for every word. We used this approach.

In addition, we consider it to be important that payment for the real effort task made is in cash and is carried out directly afterwards and before the actual experiment (the selection of one of the two lotteries). Willingness to spend is noticeably reduced if payment is made in cash in comparison to credit or debit cards (see, for example, Prelic and Semester, 2001; Runnemark et al., 2015). It has also been shown that impulsive purchase behavior is restricted when a person is handling cash (see, for example, Thomas, Kaushik and Seenivasan, 2011). From this we conclude that immediate cash payment after the real effort task leads to the

___

[8] We also chose this path in order to obtain maximum credibility with regard to an uninfluenced random process (see also Chapter 3).

subjects perceiving the amount as their own hard-earned money. In this way, the so-called house money effect[9] is probably avoided or at least considerably reduced.

We pay the subjects a show-up fee of €2. For the coding work (real effort task) the subjects earn €25, for which they require between 35 and 60 minutes. In the actual experiment the subjects earn an average of €5.56. Overall the subjects thus earn an average of €32.56. Reading the instructions for the game, answering the test questions, carrying out the coding work, deciding between piles A and B and taking a card take up between 60 and 90 Minutes. The payment they receive is therefore at an appropriate, average level. The subjects gave the impression of being very attentive and motivated.

The results were clear and are largely in line with our expectations (Table 6). In Treatment 1 (no possibility of loss), only 21 out of 53 subjects (39.62%) chose the low-risk variation (pile A). 28 subjects (52.83%) chose the risky variant (pile B). Four subjects (7.55%) were indifferent to whether they chose a card from pile A or B. In Treatment 2 (possibility of a small loss), 25 out of 52 subjects (48.08%) chose the low-risk variation (pile A). 24 subjects (46.15%) chose the risky variant (pile B). Three subjects (5.77%) were indifferent to whether they chose a card from pile A or B. In Treatment 2, the low-risk pile A was chosen more often and the risky pile B was chosen less frequently than in Treatment 1. However, Pearson's chi squared test showed this difference to be insignificant, with a p-value of 0.418 (Table 7). The null hypothesis 1 therefore has to be rejected. This confirmed our presumption (hypothesis 1) that a possibility of a small loss of €-2.50 does not have any significant influence on the risk preferences of the subjects. This result is in line with the findings of Crosetto and Filippin (2013) and those of Eckel and Grossman (2008).

Table 6: Results of the Selection Decision According to Treatments

Treat-ment	Pile A	Pile B	Decision for Pile A Number	Decision for Pile B Number	Indif-ferent Number	Decision for Pile A in %	Decision for Pile B in %	Indif-ferent in %
1	€+4/€+6	€±0/€+10	21	28	4	39.62%	52.83%	7.55%
2	€+4/€+6	€-2.5/€+12.5	25	24	3	48.08%	46.15%	5.77%
3	€+4/€+6	€-25/€+35	36	11	5	69.23%	21.15%	9.62%

Table 7: Results of Pearson's Chi Squared Test

Comparison	P-Value
Treatment 1 (no possibility of loss) versus Treatment 2 (possibility of a small loss)	0.418
Treatment 1 (no possibility of loss) versus Treatment 3 (possibility of significant loss)	0.001
Treatment 2 (low possibility of loss) versus Treatment 3 (possibility of significant loss)	0.009

[9] Thaler and Johnson (1990) showed that subjects take more risks when they have previously made a profit or if start-up capital is made available to them. This applies as long as their earlier profit or start-up capital have not been used up, and they are playing with 'house money', as it were.

In Treatment 3, however (possibility of a significant loss) a marked influence on risk preferences can be noted. Here, 36 out of 52 subjects (69.23%) chose the low-risk variation (pile A). Only eleven subjects (21.15%) chose the risky variant (pile B). Five subjects (9.62%) were indifferent to whether they chose a card from pile A or B. This is a marked difference in comparison to Treatment 1 (no possibility of loss). Pearson's chi squared test also shows this difference with a p-value of 0.001 (Table 7). Null hypothesis 2 clearly has to be rejected. Our presumption that the possibility of a higher loss (€-25) leads more often to risk-averse behavior (hypothesis 2) is thus confirmed.

The comparison between Treatment 2 (risk of a small loss) and Treatment 3 (risk of high loss) also reveals considerable differences. Pearson's chi squared test shows this difference with a p-value of 0.009 (Table 7). It is therefore clear that null hypothesis 3 also has to be rejected. Our presumption that a risk of a high loss influences the risk preference of subject considerably more than a risk of a low loss (hypothesis 3) is thus confirmed. Crosetto and Filippin (2013) had already expressed the presumption that a probability of a high loss would have an effect on the measurement of risk preferences. Our results entirely confirm this presumption.

Overall, it can be stated that taking a probability of a substantial loss into account leads to a more realistic recording of the three categories of risk preference (risk-averse, risk-neutral and risk-loving).

5 Conclusion

Experimental research on diversification behavior requires a clear differentiation between risk-averse, risk-neutral and risk-loving subjects, because decisions which can be absolutely meaningful for a risk-loving subject are completely inconceivable for a risk-averse subject and vice-versa. Robust findings in experimental research on diversification can only be obtained if it is known how to categorize the risk preferences of the subject. Differentiating between risk-neutral, risk-averse and risk-loving subjects is, however, a demanding task. The approach used by Holt and Laury (2002) has undoubtedly received the most attention. We have also used this procedure on several occasions (see, for example, Filiz et al., 2018; Gubaydullina and Spiwoks 2015). However, we also had the impression that not all subjects dedicate themselves to the task with the necessary concentration, and in view of its complexity ultimately make spontaneous decisions which are not well-thought out (for similar observations see Jacobson and Petrie, 2009; Charness and Viceisza, 2011).

The approach used by Eckel and Grossman (2008) is significantly simpler and clearer, and that deployed by Crosetto and Filippin (2013) even more so. However, all three procedures exhibit the weakness that in certain situations it is not possible to differentiate in an unambiguous and reliable way between risk-averse, risk-neutral and risk-loving subjects. In addition, in these three approaches the influence of loss aversion on risk preference is not taken into consideration, or not sufficiently.

In the form of our Treatment 3 (probability of a substantial loss) we are proposing a new approach to discriminate between risk-averse, risk-neutral and risk-loving subjects which is (1)

extremely simple and clear, and which (2) permits the clear assignment of subjects to the three categories of risk preference, and (3) takes the influence of loss aversion on risk preference into account in an appropriate way.

Acknowledgments

References

Agnew, J. / Balduzzi, P. / Sunden, A. (2003): Portfolio Choice and Trading in a Large 401(k) Plan, American Economic Review, 93(1), 193-215.

Anderson, A. (2013): Trading and Under-Diversification, Review of Finance, 17(5), 1-43.

Baltussen, G. / Post, G. T. (2011): Irrational Diversification: An Examination of Individual Portfolio Choice, Journal of Financial and Quantitative Analysis, 46(5), 1463-1491.

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

Bechara, A. / Damasio, A. R. / Damasio, H. / Anderson, S. W. (1994): Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex, Cognition, 50, 7-15.

Benartzi, S. (2001): Excessive Extrapolation and the Allocation of 401(K), Accounts to Company Stock, Journal of Finance, 56(5), 1747-1764.

Benartzi, S. / Thaler, R. H. (2001): Naïve Diversification Strategies in Defined Contribution Saving Plans, American Economic Review, 91(1), 79-98.

Benndorf, V. / Rau, H. A. / Sölch, C. (2014): Minimizing Learning Behavior in Experiments with Repeated Real-Effort Tasks, SSRN, 2503029.

Blume, M. E. / Friend, I. (1975): The Asset Structure of Individual Portfolios and some Implications for Utility Functions, Journal of Finance, 30(2), 585-603.

Bode, M. / van Echelpoel, A. / Sievi, C. R. (1994): Multinationale Diversifikation: Viel Zitiert, Kaum Befolgt, Die Bank, 94(4), 202-206.

Charness, G. / Eckel, C. / Gneezy, U. (2018): Complexity in Risk Elicitation may the Conclusions: A Demonstration Using Gender Differences, Journal of Risk and Uncertainty, 56, 1-17.

Charness, G. / Gneezy, U. / Imas, A. (2013): Experimental Methods: Eliciting Risk Preferences, Journal of Economic Behavior and Organization, 87, 43-51.

Charness, G. / Viceisza, A. (2011): Comprehension and Risk Elicitation in the Field: Evidence from Rural Senegal, IFPRI Discussion Papers, 1135.

Choi, J. J. / Laibson, D. / Madrian, B. C. / Metrick, A. (2004): For Better or for Worse: Default Effects and 401(K) Savings Behavior, Wise, D. A. (Ed.), Perspectives in the Economics of Aging, University of Chicago Press, 81-121.

Crosetto, P. / Filippin, A. (2013): The 'Bomb' Risk Elicitation Task, Journal of Risk and Uncertainty, 47(1), 31-65.

Dave, C. / Eckel, C. C. / Johnson, C. A. / Rojas, C. (2010): Eliciting Risk Preferences: When is Simple Better? Journal of Risk and Uncertainty, 41(3), 219-243.

Dimmock, S. G. / Kouwenberg, R. / Mitchell, O. S. / Peijnenburg, K. (2016): Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence, Journal of Financial Economics, 119, 59-577.

Eckel, C. C. / Grossman, P. J. (2002): Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk, Evolution and Human Behavior, 23(4), 281-295.

Eckel, C. C. / Grossman, P. J. (2008): Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices, Journal of Economic Behavior and Organization, 68(1), 1-17.

Erkal, N. / Gangadharan, L. / Nikiforakis, N. (2011): Relative Earnings and Giving in a Real-Effort Experiment, American Economic Review, 101(7), 3330-3348.

Eyster, E. / Weizsäcker, G. (2011): Correlation Neglect in Financial Decision Making, DIW Discussion Paper, 1104.

Fehr, E. / Goette, L. (2007): Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment, American Economic Review, 97(1), 298-317.

Fernandes, D. (2013): The 1/N Rule Revisited: Heterogeneity in the Naïve Diversification Bias, International Journal of Research in Marketing, 30(3), 310-313.

Filiz, I. / Nahmer, T. / Spiwoks, M. / Bizer, K. (2018): Portfolio Diversification: The Influence of Herding, Status-Quo Bias and the Gambler's Fallacy, Financial Markets and Portfolio Management, 32(2), 167-205.

Fischbacher, U. (2007): Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, Experimental Economics, 10(2), 171-178.

Gneezy, U. / Potters, J. (1997): An Experiment on Risk Taking and Evaluation Periods, Quarterly Journal of Economics, 112(2), 631-645.

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

Gubaydullina, Z. / Spiwoks, M. (2015): Correlation Neglect, Naïve Diversification, and Irrelevant Information as Stumbling Blocks for Optimal Diversification, Journal of Finance and Investment Analysis, 4(2), 1-19.

Guiso, L. / Haliassos, M. / Japelli, T. (2002): Household Portfolios, MIT Press, Cambridge.

Hedesstrom, T. M. / Svedsater, H. / Garling, T. (2006): Covariation Neglect among Novice Investors, Journal of Experimental Psychology: Applied, 12(3), 155-165.

Hibbert, A. M. / Lawrence, E. R. / Prakash, A. J. (2012): Can Diversification be Learned? Journal of Behavioral Finance, 13(1), 38-50.

Holt, C. A. / Laury, S. K. (2002): Risk Aversion and Incentive Effects, American Economic Review, 92(5), 1644-1655.

Huberman, G. / Sengmueller, P. (2004): Performance and Employer Stock in 401(K) Plans, Review of Finance, 8(3), 403-443.

Filiz, I. (2018): Overconfidence: Der Einfluss Positiver und Negativer Effekte, Sofia Diskussionsbeiträge zur Institutionenanalyse, 17(1).

Jacobsen, S. / Petrie, R. (2009): Learning from Mistakes: What Do Inconsistent Choices over Risk Tell Us? Journal of Risk and Uncertainty, 38(2), 143-158.

Kahneman, D. (2009): Thinking, Fast and Slow, London.

Kallir, I. / Sonsino, D. (2009): The Neglect of Correlation in Allocation Decisions, Southern Economic Journal 75(4), 1045-1066.

Lease, R. C. / Lewellen, W. G. / Schlarbaum, G. G. (1974): The Individual Investor: Attributes and Attitudes, Journal of Finance, 29(2), 413-433.

Lejuez, C. W. / Read, J. P. / Kahler, C. W. / Richards, J. B. / Ramsey, S. E. / Stuart, G. L. / Strong, D. R. (2002): Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (Bart), Journal of Experimental Psychology: Applied, 8(2), 75-84.

Lönnqvist, J. E. / Verkasalo, M. / Walkowitz, G. / Wichardt, P. C. (2015): Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison, Journal of Economic Behavior and Organization, 119, 254-266.

Markowitz, H. (1952): Portfolio Selection, Journal of Finance, 7(1), 77-91.

Meulbroek, L. (2005): Company Stock in Pension Plans: How Costly is it? Journal of Law and Economics, 48(2), 443-474.

Morrin, M. / Inman, J. J. / Broniarczyk, S. M. / Nenkov, G. Y. / Reuter, J. (2012): Investing for Retirement: The Moderating Effect of Fund Assortment Size on the 1/N Heuristic, Journal of Marketing Research, 49(4), 537-550.

Mukherjee, S. / Pammi, V. S. C. / Sahay, A. / Srinivasan, N. (2017): Is Loss-Aversion Magnitude-Dependent? Measuring Prospective Affective Judgments Regarding Gains and Losses, Judgment and Decision Making, 12(1), 81-89.

Polkovnichenko, V. (2005): Household Portfolio Diversification: A Case for Rank-Dependent Preferences, Review of Financial Studies, 18(4), 1467-1502.

Prelec, D. / Simester, D. (2001): Always Leave Home Without it: A Further Investigation of the Credit-Card Effect on Willingness to Pay, Marketing Letters, 12(1), 5-12.

Rabin, M. (2000): Risk Aversion and Expected-Utility Theory: A Calibration Theorem, Econometrica, 68(5), 1281-1293.

Rieger, M. O. (2012): Why Do Investors Buy Bad Financial Products? Probability Misestimation and Preferences in Financial Investment Decision, Journal of Behavioral Finance, 13(2), 108-118.

Runnemark, E. / Hedman, J. / Xiao, X. (2015): Do Consumers Pay More Using Debit Cards Than Cash?, Electronic Commerce Research and Applications, 14(5), 285-291.

Schupp, J. / Wagner, G. G. (2002): Maintenance of and Innovation in Long-Term Panel Studies, the Case of the German Socio-Economic Panel (Gsoep), DIW Discussion Paper, 276.

Thaler, R. H. / Johnson, E. J. (1990): Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice, Management Science, 36(6), 643-660.

Thomas, M. / Kaushik Desai, K. / Seenivasan, S. (2011): How Credit Card Payments Increase Unhealthy Food Purchases: Visceral Regulation of Vices, Journal of Consumer Research, 38(1), 126-139.

Tom, S. M. / Fox, C. R. / Trepel, C. / Poldrack, A. R. (2007): The Neural Basis of Loss Aversion in Decision-Making under Risk, Science, 315(5811), 515-518.

Wagner, G. G. / Burkhauser, R. V. / Behringer, F. (1993): The English Language Public Use File of the German Socio-Economic Panel, Journal of Human Resources, 28(2), 429-433.

Weber, E. U. / Blais, A. R. / Betz, N. E. (2002): A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors, Journal of Behavioral Decision Making, 15(4), 263-290.

Weber, E. U. / Siebenmorgen, N. / Weber, M. (2005): Communication Asset Risk: How Name Recognition and the Format of Historic Volatility Information Affect Risk Perception and Investment Decisions, Risk Analysis, 25(3), 597-609.

Zuckermann, M. (1979): Sensation Seeking: Beyond the Optimal Level of Arousal, John Wiley and Sons Inc, New York.

Zuckermann, M. (1994): Behavioral Expressions and Biosocial Bases of Sensation Seeking, Cambridge University Press, Cambridge.

Appendix: Instructions for the Game, Test Questions and Selected Screen-shots

The main instructions are from the study by Benndorf, Rau and Solch (2014).

<u>The Task</u>

In this game, you can earn money by completing a task. The task consists of **coding 175 words** into numbers. For every correctly coded word you receive credit of **14 cents.** In total you can earn up to **€25**. In the task, **three upper case letters** correspond to one word. Every upper-case letter must be assigned to a number. The coding for this can be found in the table below. Please take a look at the photo on the screen:

In this example, the participant has already coded two words correctly. Now the three upper case letters have to be coded: **V, B and U**. The solution is provided in the table:

- V is 398. (see the number entered above by the participant)

- B is 463.

- U is 575.

To enter please click on the blue box below the first upper-case letter.

When all three figures have been entered please click on OK with the mouse.

The computer then checks whether <u>**ALL**</u> upper-case letters have been correctly coded into figures, i.e. whether all three figures were entered correctly. Only then is the word considered to be correct. If the wrong number is entered, the computer points this out (in red letters) after pressing the OK button. The current word remains on the screen until the correct number is entered. However, your previous entries (in the three fields under the upper-case letters)

are all deleted. The table remains the same, i.e. the numbers assigned to the letters remain identical. In the same way, the position of the upper-case letters in the table does not change.

When the correct number is entered you receive the next randomly drawn word (again consisting of three upper-case letters). The table continues to be randomly 're-shuffled': New three-digit figures are randomly selected and entered into the table as new mappings for the upper-case letters. The position of the upper-case letters in the table is randomly rearranged. Please note that all 26 upper-case letters of the German alphabet are used.

Following on from this task you will take part in a lottery. You will be given detailed information about the lottery when the time comes.

Payment

- Basic payment of **€2.**
- For every correctly coded word you receive credit of **14 cents**. In total you can earn up to **€25** (175 x 14.2857 cents).
- The money you might win in the lottery is added to this.

Important Information

- Please remain quiet during the game!
- Do not look at your neighbor's screen!
- No aids are permitted (calculators, smartphones etc.) All electronic devices must be switched off!

Test Questions (Treatment 1)

Test question 1: How much is the minimum and maximum payment when you choose pile A?

    a. The minimum payment is €+6 and the maximum payment is €+35.

    b. The minimum payment is €+4 and the maximum payment is €+6. (correct)

    c. The minimum payment is €+4 and the maximum payment is €+35.

Test question 2: How much is the minimum and maximum payment when you choose pile B?

    a. The minimum payment is €±0 and the maximum payment is €+35.

    b. The minimum payment is €+4 and the maximum payment is €+10.

    c. The minimum payment is €±0 and the maximum payment is €+10. (correct)

Test question 3: How many different piles of cards are there?

    a. 1

    b. 2 (correct)

    c. 3

Test question 4: How high is the probability of occurrence of the best and worst possible results in the lottery?

a. 100%

b. 0%

c. 50% (correct)

Screenshot 1: Real Effort Task (Reconstructed in Order to Improve Readability)

Coding

Out of **175** words to be coded you are currently encoding word number **1**

Word: Y   E   S

Code: [   ]   [   ]   [   ]

T	C	Y	R	K	Q	V	Z	N	X	D	A	M	U	S	P	W	G	J	B	F	I	Q	H	L	E
735	105	494	343	737	691	825	865	771	587	204	606	362	769	155	345	287	288	165	977	208	200	177	291	791	660

O.K.

Figure A-1: Real Effort Task

Lottery

You can now take a card as part of a lottery. There are two piles of cards to choose from (pile A and pile B). The cards are drawn by hand.

There are four cards in pile A. Two cards lead to a payment of **€+4** and two cards lead to a payment of **€+6.**

There are four cards in pile B. Two cards lead to a payment of **€±0** and two cards lead to a payment of **€+10.**

Pile A	Pile B
(consists of four cards)	**(consists of four cards)**

Best possible event: €+6 (Probability: 50%)
Worst possible event: €+4 (Probability: 50%)
Expected value: €+5
Low-risk (results fluctuate slightly around the expected value)

Best possible event: €+10 (Probability: 50%)
Worst possible event: €±0 (Probability: 50%)
Expected value: €+5
High-risk (results fluctuate considerably around the expected value)

Please answer the following test questions about the lottery:

Test question 1: How much is the minimum and maximum payment when you choose pile A?

    ◯ The minimum payment is €+6 and the maximum payment is €+35.

    ◯ The minimum payment is €+4 and the maximum payment is €+6.

    ◯ The minimum payment is €+4 and the maximum payment is €+35.

Test question 2: How much is the minimum and maximum payment when you choose pile B?

    ◯ The minimum payment is €+/-0 and the maximum payment is €+35.

    ◯ The minimum payment is €+4 and the maximum payment is €+10.

    ◯ The minimum payment is €+/-0 and the maximum payment is €+10.

Test question 3: How many different piles of cards are there?

    O 1

    O 2

    O 3

Test question 4: How high is the probability of occurrence of the best and worst possible results in the lottery?

    O 100%

    O 0%

    O 50%

<div align="center">

O.K.

</div>

Figure A-2: Instructions and Test Questions on the Lottery in Treatment 1
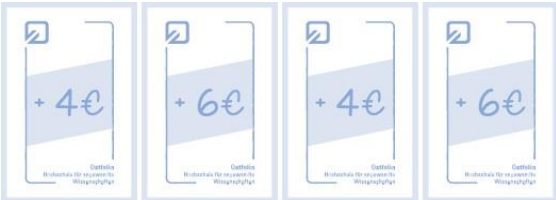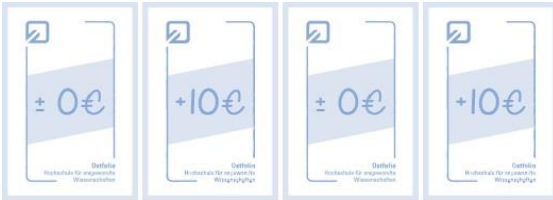
Screenshot 3: Decision-Making Options on the Lottery in Treatment 1 (Reconstructed in Order to Improve Readability)

Lottery

You can now take a card as part of a lottery. There are two piles of cards to choose from (pile A and pile B). The cards are drawn by hand.

There are four cards in pile A. Two cards lead to a payment of **€+4** and two cards lead to a payment of **€+6.**

There are four cards in pile B. Two cards lead to a payment of **€±0** and two cards lead to a payment of **€+10**.

<div align="center">

Pile A

(consists of four cards)

</div>

Best possible event: €+6 (Probability: 50%)
Worst possible event: €+4 (Probability: 50%)
Expected value: €+5
Low-risk (results fluctuate slightly around the expected value)

<div align="center">

Pile B

(consists of four cards)

</div>

Best possible event: €+10 (Probability: 50%)
Worst possible event: €±0 (Probability: 50%)
Expected value: €+5
High-risk (results fluctuate considerably around the expected value)

You now have to decide which pile you want to take a card from.

Make your selection now. Click on one of the three alternatives.

Ο I would like to take a card from pile A.

Ο I would like to take a card from pile B.

Ο I would like to take a card, but I don't mind which pile I take one from.

<div align="center">

O.K.

</div>

Figure A-3: Decision-Making Options on the Lottery in Treatment 1

Chapter 5

HERD BEHAVIOR AND MOOD
AN EXPERIMENTAL STUDY ON THE
FORECASTING OF SHARE PRICES

with Ibrahim Filiz and Markus Spiwoks
contribution Thomas Nahmer: 45%

Abstract:

This experimental study examines the influence of emotions on the tendency towards herd behavior. The subjects forecast share prices, and while doing so they are offered the chance to orientate themselves towards other subjects and to possibly exhibit herding behavior. In a between-subjects design, three treatments are used (neutral, positive and negative mood). Mood is influenced by means of film excerpts which are shown to the subjects. It is shown that mood really does have an influence on the tendency towards herding behavior. A neutral mood in particular favors a tendency towards herd behavior.

1 Introduction

Herding usually refers to convergent social behavior. A common orientation of thoughts or behavior occurs not as a result of centralized coordination, but via the interaction of individuals. For centuries now there have been reports of events which correspond to this understanding of herd behavior. As early as 1374, for example, there were reports of people entering a trance-like state while dancing, which led to mass hysteria and even deaths (Waller, 2009). Other forms of social contagion can occur, for example, during social unrest, due to political opinions or in modern hooliganism (Le Bon, 1896; Russel, 2004). Raafat, Chater and Frith (2009) illustrated a wide range of historical examples for the occurrence of herd behavior.

In economics, herding takes place when subjects imitate the behavior of other subjects, or make economic decisions on the basis of the views, assessments or actions of others. A subject may have good reasons to exhibit herd behavior. For example, financial market analysts might align themselves with the majority opinion because in this way they can best avoid the danger of damage to their reputations. Institutional investors might imitate the investment behavior of others because they think that in this way they can increase their earnings. However, many investors copy the behavior of other investors in an ill-advised and completely irrational way - for example when they get carried along by what is generally thought to be a favorable stock-market mood, or when they panic in the face of falling share prices. Sometimes, social conventions or customs also play a role in the occurrence of herd behavior (Spyrou, 2013).

Adam Smith (1759) and Charles Mackay (1841) were among the first to highlight the significance of herding for the explanation of economic trends. Keynes (1936) also pointed out herding behavior among financial market participants and presented two explanations for it - reputational herding and investigative herding. Scharfstein and Stein (1990) took up these thoughts and thus initiated an intensive discourse which has continued over the past three decades.

Keynes (1936) assumed that financial market analysts tended towards herding because they had little to gain by having a highly individual opinion, and at the same time would be risking a considerable loss of reputation if it went wrong. However, if they aligned themselves with the majority opinion, they might be able to enhance their reputations slightly without taking the risk of endangering them. Financial market analysts who want to keep their well-paid jobs are thus behaving rationally when they always follow the opinion of the majority. In addition, Keynes (1936) assumed that financial market participants with a short-term orientation also tended towards investigative herding. Those who bet on short-term movements in the financial markets should not rely on information which will only be noticed months later by other capital market actors. They would be wiser to orientate themselves towards that which others consider to be relevant in the short term.

The reputational herding and investigative herding approaches have been examined in a large number of studies. Cote and Sanders (1997) came to the conclusion that concerns about reputations and the particular credibility of consensus forecasts favor herd behavior, which in turn leads to a worsening of the quality of forecasts. Bedke, Bizer and Spiwoks (2009) as well as Meub et al. (2015) confirmed via experiments that concerns about reputations are a decisive factor in the formation of rational herding behavior.

Some studies found a negative correlation between professional experience in years and the tendency towards herding behavior (Hong, Kubik and Solomon, 2000; Lamont, 2002; Clement and Tse, 2005; Krishnan, Lim and Zhou, 2005). Inexperienced analysts who diverge considerably from the majority opinion with their forecasts are at greater risk of being made redundant than experienced analysts. Ashiya and Doi (2001) on the other hand found no indications that

the experience of analysts had any significance on the occurrence of herding behavior. Some studies show that analysts who were particularly successful in the past are less likely to exhibit herding behavior (Stickel, 1990; Cote and Sanders, 1997; Graham, 1999). Bhalla (2012) on the other hand, found no indications that analysts make independent forecasts. Spiwoks, Bizer and Hein (2008b) considered how coordinated behavior among financial market analysts can come about. They found that an orientation towards the naive forecast ensures that analysts lastingly remain within the protective environment of the herd. Others have also shown that the more difficult the relevant tasks the earlier herding behavior occurs (Olsen, 1996; Kim and Pantzalis, 2003).

Banerjee (2010) as well as Bikhchandani, Hirshleifer and Welch (1992) showed that herding behavior can even occur when subjects act rationally and make a genuine effort to take meaningful decisions (informational cascades). In this case the subjects try - on the basis of observable decisions made by others - to deduce the latter's private signals. All information (a priori probability, one's own private signal and the private signals of others) is then used to maximize the probability of success of one's own decision. This can lead to herding behavior. Numerous experimental studies have found indications for the fact that informational cascades can really occur (Anderson and Holt, 1997; Hung and Plott, 2001; Celen and Kariv, 2004). A large number of variations of experiments have essentially confirmed these results. In laboratory experiments, informational cascades have also occurred in the case of delayed decision-making so that subjects could benefit from observing others' actions (Sgroi, 2003), or when there was the opportunity to purchase additional information (Kübler and Weizsäcker, 2004; Kraemer, Nöth and Weber, 2006); in the case of decisions which were not strictly sequential (Orlean, 1995), and in the case of different signal qualities (Sasaki, 2005; Sumpter et al. 2012). Other experimental studies have expressed major doubts about how realistic informational cascades are (Spiwoks, Bizer and Hein, 2008a; Çelen and Kariv, 2004; Nöth and Weber, 2003; Huck and Oechssler, 2000). Some subjects assign an inappropriately high level of significance to their own private signal, while others falsely interpret the behavior of those who have made previous decisions. Some subjects orientate themselves solely towards a priori probability, while others orientate themselves exclusively towards the majority of previous decision-makers. Only a small number of subjects are able to apply the Bayesian rule, and many subjects simply follow their gut feeling. Only very few grasp all the relevant information and make rational deductions about the probability of success of the different alternatives. This frequently leads to overlaps in behavior which look like informational cascades but have nothing in common with the thoughts at the core of the concept.

Devenow and Welch (1996) were the first to make a clear differentiation between rational herding (reputational herding, investigative herding and informational cascades) on the one hand and irrational herding behavior on the other. Irrational herding behavior often has a cultural background (social conventions) or is due to a lack of control over one's impulses and can only be explained psychologically (Baddeley, Curtis and Wood, 2004; Baddeley, 2010). Shleifer and Summers (1990) emphasized the irrational herding behavior of noise traders, for example. Their actions are frequently characterized by pseudo-signals such as recommendations by financial gurus (Black, 1986). Misinterpretations of the actions of third parties can also trigger irrational herding. For example, the purchase of a share can have the purpose of merely restoring the original weighting of a portfolio (rebalancing). Other investors might, however, interpret this purchase as an indication of the particular attractiveness of the share in question and therefore buy it too. Presumptions about the motives of other investors can form a distraction from more important information and lead to sub-optimal decisions

(Simonsohn and Ariely, 2008). Roider and Voskort (2016) interpret occurrences of herding behavior as unreflected mutual imitation.

Bikhchandani and Sharma (2000) pointed out that similar behavior should not be equated with herding. If many subjects do the same thing without taking note of each other or reacting to each other in any way, this is referred to as spurious herding. Galariotis, Rong and Spyrou (2015) presented empirical indications of spurious herding. A tendency towards non-conformism can, however, lead to individual subjects consciously leaving a herd. For this phenomenon, known as anti-herding, several findings have been presented (Zitzewitz, 2001; Laux and Probst, 2004; Bernhardt, Campello and Kutsoati, 2006; Naujoks et al., 2009; Pierdzioch and Rülke, 2012; Pierdzioch, Rülke and Stadtmann, 2013).

Table 1: Overview of herding behavior

intentional herding				spurious herding
rational herding			irrational herding	
reputational herding	investigative herding	informational cascades		

It is now considered certain that mood has an influence on economic decision-making (for an overview see, for example, George and Dane, 2016; Lerner et al., 2015; Vohs, Baumeister and Loewenstein, 2007; Baker and Wurgler, 2007; Baumeister et al., 2007; Pham, 2007; Shiv et al., 2005; Nofsinger, 2005; Lucey and Dowling, 2005; Watson and Vaidya, 2003; Daniel, Hirshleifer and Teoh, 2002; Hirshleifer, 2001; Loewenstein et al., 2001; Isen, 2000; Loewenstein, 2000; Schwarz, 2000; Elster, 1998; Bless, Schwarz and Kemmelmeier, 1996; Elster, 1996; Johnson and Tversky, 1983).

For example, the effects of sunshine, rain, cloud cover, wind strength, storms and other meteorological factors on market returns at stock exchanges worldwide have been thoroughly investigated (Kim, 2017; Kaustia and Rantapuska, 2016; Apergis, Gabrielsen and Smales, 2016; Bassi, Colacito and Fulghieri, 2013; Lu and Chou, 2012; Mirza et al., 2012; Floros, 2011; Symeonidis, Daskalakis and Markellos, 2010; Kang et al., 2010; Shu and Hung, 2009; Chang et al., 2008; Keef and Roush, 2007; Chang et al., 2006; Dowling and Lucey, 2005; Cao and Wei, 2005; Tufan and Hamarat, 2004; Krivelyova and Robotti, 2003; Hirshleifer and Shumway, 2003; Kamstra, Kramer and Levi, 2003; Pardo and Enric, 2002; Krämer and Runde, 1997; Saunders, 1993). While doing so, attempts were also made to create a connection between the weather and the mood of capital market protagonists: Hirshleifer and Shumway (2003) showed that stock market returns on days when the sun shines in the morning were higher on average than on days with bad weather. This result was explained by sunshine favoring a positive atmosphere among investors. Kamstra, Kramer and Levi (2003) established that stock market returns varied according to the length of the day, which has been interpreted in a similar way to the results of Hirshleifer and Shumway (2003). Kaustia and Rantapuska (2016) carried out a similar study - however, they only observed a weak correlation between the effect of the length of a respective day, the weather and investment decisions.

Experimental economic research is increasingly interested in the question of which influence positive and negative emotions have on investment decisions. Grable and Roszkowski (2008), for example, showed in an experimental study that subjects whose positive emotions predominate were willing to take greater financial risks, while Kuhnen and Knutson (2011) carried out

experiments to establish how different moods affected investment decisions. This showed that subjects tended to avoid risks when their mood is negative, while subjects whose mood was positive were willing to take more risks. According to this, subjects whose emotions are positive are more optimistic in relation to their investment decisions. Colasante, Marini and Russo (2017), however, came to the opposite conclusion. Their subjects were more risk-averse when under the influence of positive and negative emotions than those in a neutral mood. Kaplanski et al. (2015) showed that the mood of investors had an influence on their expectations in terms of returns, and also on their perception of risk. The happier the subjects were, the greater were their expectations in terms of their returns, and the lower the presumed risk of stock market investments. Experiments carried out by Lee and Andrade (2014) showed that negative affects increase risk aversion in investment decisions. An overview of the studies dealing with the effects of mood on risk tolerance can be viewed in Lane (2017) and Duxbury (2015). Lahav and Meer (2012) as well as Andrade, Odean and Lin (2016) used experiments to examine the effect of emotions on speculative bubbles. They established that speculative bubbles were larger in the case of positive affects than with negative affects. Breaban and Noussair (2018) followed a similar approach, though their findings were not as clear-cut. Gavriilidis, Kallinterakis and Tsalavoutas (2016) established that the positive atmosphere during Ramadan leads to an increase in herding behavior on the capital markets.

By now there are thus a range of findings showing that the mood of subjects can influence their economic decisions. However, until now no-one has conducted experiments on whether different moods also have an effect on the likelihood of herding behavior. Precisely this research topic has now been addressed by this study. The task facing the subjects here was to forecast future share prices.

2 Design of the experiment and hypotheses

Forecasting future share prices is a challenge which active portfolio managers have to face on a daily basis. However, the percentage of successful share price forecasts is modest. The neo-classical theory of economics assumes strong or at least semi-strong informational efficiency. In an environment of this kind, forecasting future capital market trends is impossible. However, the conditions for informationally efficient capital markets - namely rational and fully informed subjects with a uniform formation of expectations - are now clearly considered to be non-existent. Successful share price forecasts thus appear to be possible but are seldom achieved in practice. The task of behavioral finance is to determine the cause of this apparent contradiction. Herding behavior can lead to stock market analysts using the available information in a sub-optimal way.

We orientated ourselves towards the approach used by Cote and Sanders (1997) to measure herding. The subjects have to make an initial estimate of a future share price. Subsequently they are informed about the average of the forecasts of all subjects (consensus forecast). Then they are allowed to change their original forecast once. If the subjects change their forecast in the direction of the consensus forecast, this is considered to be herding. If they change their forecast in the opposite direction, this has to be viewed as anti-herding. If the subjects do not change their forecast at all, neither herding nor anti-herding are present.

With regard to the making of forecasts, we roughly orientated ourselves towards the studies by Meub et al. (2015) and that of Becker, Leitner and Leopold-Wildburger (2009). This

approach enables subjects to determine future share price levels via four fundamental influencing factors which have a constant effect on the share price. An error-free forecast is made more difficult by a random influence which leads to a situation where even subjects who act rationally only manage to successfully forecast an average of 40% of share prices. The more the subjects deviate from a rational strategy, the lower the success rate of their forecasts. As a reward is given for every successful forecast, subjects have a financial incentive to make rational forecasts. The issue of a consensus forecast makes it possible for subjects to change their own forecast, and it can then be seen to what extent herding behavior leads to subjects diverging from a rational strategy.

In order to examine the influence of mood on the occurrence of herding, we formed three treatments (positive, negative and neutral). The influencing of moods in the experiment was carried out by showing film excerpts, which is considered to be a tried and tested technique (Westermann et al., 1996; Allwood, Granhag and Jonsson, 2002; Kirchsteiger, Rigottii and Rustichini, 2006; Rottenberg, Ray and Gross, 2007; Ifcher and Zarghamee, 2014; Oswald, Proto and Sgroi, 2015). Whether the subjects' mood has been successfully influenced is assessed with manipulation checks.

Subjects make their forecasts as follows: The share price in € at the point in time t (K_t) is determined by four influencing factors (the fundamental data A_t, B_t, C_t and D_t) and a random influence (ε_t). The fundamental data is provided before each forecast is made. The subjects are also aware of the specific influence the fundamental data has on the share price.

(1) $$K_t = 2 \cdot A_t + 3 \cdot B_t - 1 \cdot C_t + 4 \cdot D_t + \varepsilon_t$$

The random influence (ε_t) moves within a range of €-40 and €+40. The probability of the events €-40, €-30, €-20, €+20, €+30 and €+40 is 10% each. The probability of the events €-10 and €+10 is 12%. The event €±0 has a probability of occurrence of 16% (Figure 1).

Fig. 1: The probability distribution of the random influence ε_t on the formation of the share price in €
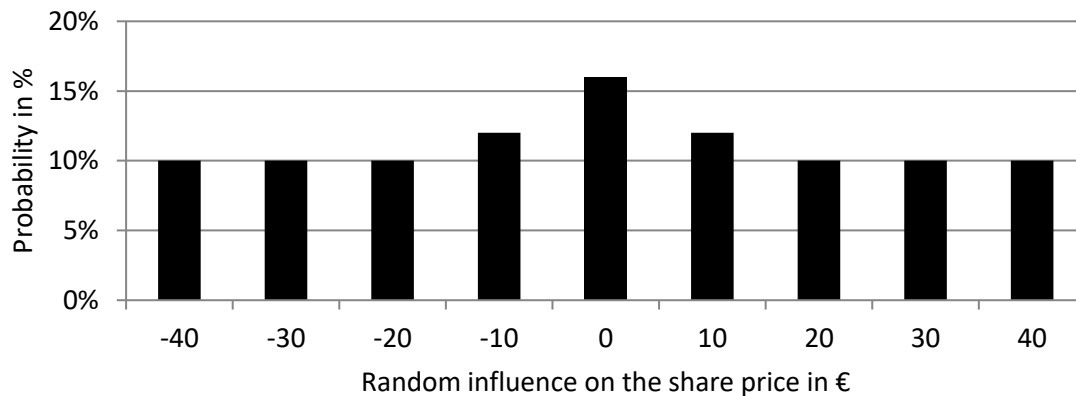
The random influence for each period is determined by means of a lottery. As an example, here is how the share price was determined for Period 1. The subjects were informed that A_1 = 32, B_1 = 26, C_1 = 22 and D_1 = 25.

Hence:

$$K_1 = 2 \cdot 32 + 3 \cdot 26 - 1 \cdot 22 + 4 \cdot 25 + \varepsilon_1 = 220 + \varepsilon_1$$

The expected value for ϵ_t is zero. In addition, €±0 is also the most frequently occurring random event (16%). A forecast is considered successful and is rewarded accordingly when it does not diverge more than €10 from the actual share price. Rational subjects will therefore, in accordance with the expected value of the price, make a forecast of €220 (in the example of Period 1). In 40% of all cases their forecasts will be correct, because the probability that the price is not more than €10 above or below the expected value is 12% + 16% + 12% = 40%. If one made a forecast of €230, the probability of success would only be 38% (16% + 12% + 10%). And if one made a forecast of €240, the probability of success would fall even further to 32% (12% + 10% + 10%). A forecast of €280 would have a probability of success of 0%, because ϵ_t moves between €-40 and €+40 and the tolerance between the actual share price and a successful forecast is only €10.

The subjects play five rounds - in other words, they are supposed to make five preliminary and five final forecasts. Each successful forecast is rewarded with €10. The show-up fee is €5.

The experiment is carried out as follows: before the beginning of the experiment an initial mood test is carried out in order to recognize any possible distortions of the research results due to an unusual mood beforehand. Then the subjects receive an instruction sheet in which the rules and requirements of the game are explained (see Appendix). Subsequently they have to answer test questions to ensure that they have understood how the game works (see Appendix). Then the subjects receive a history of the share price during the past ten periods as a line chart as well as an overview of the random events of the past ten periods as a bar chart (see Appendix).

This is followed by the first film excerpt and a subsequent manipulation check. Then the fundamental data for the current period is announced. The subjects are allowed to use a pocket calculator to determine the expected value. After this they have to make their preliminary forecast. In the next step another film excerpt is shown and a second manipulation test is carried out. Then the consensus forecast is made available and the subjects are asked whether they want to change their forecasts. Finally, the actual share price is announced and the share price history chart is completed. This procedure is repeated a total of five times. At the end the subjects' forecasting performance is determined and their performance-based payment and the show-up fee are paid.

The three treatments differ with regard to the film excerpts shown: in the treatment "positive", film excerpts are shown which usually create a positive mood. In the treatment "negative", film excerpts are shown which usually create a negative mood. And in the treatment "neutral", film excerpts are shown which normally create a neutral mood.

Numerous findings from behavioral economics show that subjects by no means always make rational decisions. We therefore expect our subjects to also make forecasts which do not correspond to the expected value of the share price. The first hypothesis is therefore as follows:

Hypothesis 1: The subjects will make forecasts which do not correspond to the expected value of the share price.

The null hypothesis to be examined is therefore:

Null hypothesis 1: The forecasts always correspond to the expected value of the share price.

On the basis of numerous findings on the occurrence of herding among capital market participants, we expect herding behavior to also appear in our experiment.

The second hypothesis is therefore as follows:

Hypothesis 2: After the announcement of the consensus forecast, adjustments will be made to the forecasts which align them closer to this consensus forecast (herding behavior).

The second null hypothesis to be examined is therefore:

Null hypothesis 2: After the announcement of the consensus forecast, no adjustments will be made to the forecasts which align them closer to this consensus forecast.

In addition, we expect that herding behavior will also be evident in the standard deviation of the forecasts before and after the announcement of the consensus forecast.

Our third hypothesis is therefore as follows:

Hypothesis 3: the standard deviation of the forecasts before the announcement of the consensus forecast (preliminary forecasts) will be higher than the standard deviation of the forecasts after the announcement of the consensus forecast (final forecasts).

The third null hypothesis to be examined is therefore:

Null hypothesis 3: the standard deviation of the forecasts before the announcement of the consensus forecast (preliminary forecasts) will not be higher than the standard deviation of the forecasts after the announcement of the consensus forecast (final forecasts).

Against the background of previous research findings on the effect of emotions, we expect that a positive or negative mood will lead to a greater deviation from the rational forecast (forecast = expected value of the share price) than a neutral mood.

The fourth hypothesis is therefore:

Hypothesis 4: In the treatment "neutral", more forecasts are made which correspond to the expected value of the share price than in the treatments "positive" and "negative".

The fourth null hypothesis to be examined is therefore:

Null hypothesis 4: In the treatment "neutral", the same number or fewer forecasts are made which correspond to the expected value of the share price than in the treatments "positive" and "negative".

In addition, we expect that subjects in a positive or negative mood will exhibit a different tendency towards herding than subjects in a neutral mood.

The fifth hypothesis is therefore:

Hypothesis 5: In the treatments "positive" and "negative", the frequency with which forecasts are corrected in the direction of the consensus forecast is different than in the treatment "neutral".

The fifth null hypothesis to be examined is therefore:

Null hypothesis 5: In the treatments "positive" and "negative", the frequency with which forecasts are corrected in the direction of the consensus forecast is the same as in the treatment "neutral".

3 Results

The experiment was carried out between 14 December and 20 December 2018 in the Ostfalia Laboratory for Experimental Economic Research (OLEW) at the Ostfalia University of Applied Sciences in Wolfsburg, Germany. A total of 181 subjects took part in the experiment. Of these, 60 played the treatment "neutral" (neutral mood), 63 played the treatment "positive" (positive mood) and 58 the treatment "negative" (negative mood).

The subjects are students at the Ostfalia University of Applied Sciences in Wolfsburg. 92 subjects (50.8%) study at the Faculty of Business, 14 subjects (7.7%) at the Faculty of Health Care, and 73 (40.3%) at the Faculty of Vehicle Technology. A total of 20 sessions were carried out.

The experiment was programmed in z-tree (Fischbacher, 2007). In the Ostfalia Laboratory for Experimental Economic Research (OLEW) there are a total of 12 computer workplaces, which makes it possible to completely separate the subjects from each other. The experiments are constantly monitored by a master of ceremonies so that communication between the subjects and the use of prohibited aids (such as smartphones) can be ruled out.

Overall the subjects received an average payment of €23.20. The highest payment was €45 and the lowest was €5. Taking part in the experiment took an average of 50 minutes, so the payment can be considered to be very attractive. The subjects gave the impression of being highly concentrated and motivated.

The influencing of the subjects' mood via the film excerpts was successful (Figure 2). The average mood in the treatment "negative" exhibited a value of 3.21, while it was 5.64 in the treatment "neutral" and 6.89 in the treatment "positive".

Fig. 2: Successful influencing of mood in the three treatments
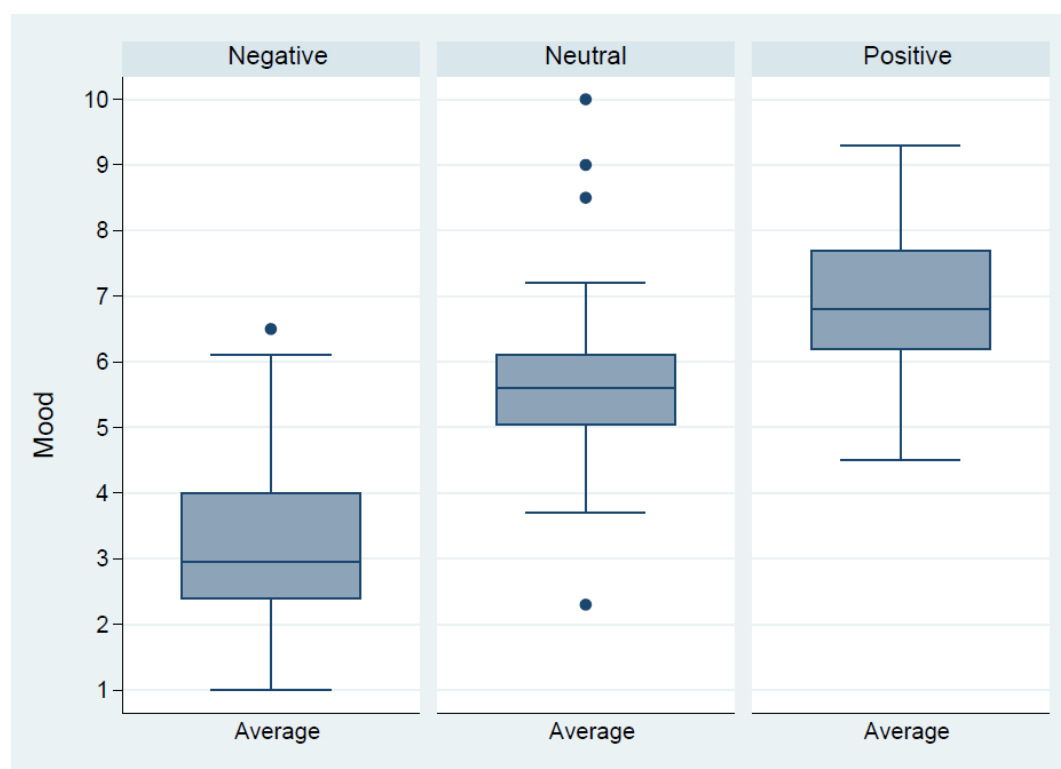
In the Mann-Whitney U test the differences in the moods of the subjects were shown to be highly significant. The treatment "negative" differed significantly from the treatment "neutral" (z = -7.838, p = 0.000). The treatment "positive" also differed significantly from the treatment "neutral" (z = -6.263, p = 0.000). Accordingly, the treatment "negative" also differed significantly from the treatment "positive" (z = -9.152, p = 0.000). In this way we established that the film excerpts achieved the intended influence on the subjects' mood.

For their forecasts, rational utility maximisers always use the expected value, which results from the four fundamental data items and the random influence. However, only just under a third (32.41%) of the preliminary forecasts and also only just below a third of the final forecasts (30.97%) corresponded to this expected value (Tab. 2). Over two thirds of all forecasts were thus not based on rational considerations, and in this way the subjects reduced the expected value of their payment. Not even one in ten subjects (9.94%) followed a rational

course of action throughout the experiment (Tab. 2). The average absolute deviation of the forecasts from the expected value was €18.26. In the t-test (one-sample mean comparison), the deviation from the rational strategy (average deviation of the of the forecasts from the expected value = 0) is highly significant (p = 0.000). Thus, <u>null hypothesis 1</u> clearly has to be rejected. This confirmed our assumption that the subjects do not correspond to the neo-classical concept of homo oeconomicus.

Table 2: Overview of the rational strategy

	Treatment „neutral"	Treatment „positive"	Treatment „negative"	All treatments
Rational strategy (preliminary forecast)	84 out of 298 forecasts 28.19%	108 out of 315 forecasts 34.29%	100 out of 290 forecasts 34.48%	292 out of 903 forecasts 32.41%
Rational strategy (final forecast)	77 out of 298 forecasts 25.84%	98 out of 315 forecasts 30.79%	104 of 290 forecasts 35.86%	279 out of 903 forecasts 30.97%
Subjects with a rational strategy in all rounds	3 out of 60 subjects 5.00%	8 out of 63 subjects 12.70%	7 out of 58 subjects 12.07%	18 out of 181 subjects 9.94%

Just over a third of the preliminary forecasts (35.99%) were revised after the announcement of the average of these forecasts (consensus forecast) (Tab. 3). The majority of these revisions (86.46%) were in the direction of the consensus forecast. For the subjects, it is easy to recognize the forecast with which they can maximize the expected value of their payment. With the aid of the information on the fundamental data, the forecast with the highest probability of success can be worked out very simply. There is therefore no need to react to the announcement of the consensus forecast, and there is no sensible reason for changing one's own preliminary forecast in the direction of the consensus forecast. The revision of forecasts in the direction of the consensus forecast therefore has to be interpreted as a sign of herding behavior. The average absolute adjustment of the forecasts in the direction of the consensus forecast amounted to €28.27. In the t-test (one-sample mean comparison), the adjustment towards the consensus forecast is shown to be highly significant (p = 0.000). Null hypothesis 2 thus also has to be rejected. Our assumption that herding behavior would occur was confirmed.

Table 3: Overview of herding behavior

	Treatment „neutral"	Treatment „positive"	Treatment „negative"	All treatments
Revised forecasts	125 out of 298 forecasts 41.95%	106 out of 315 forecasts 33.65%	94 out of 290 forecasts 32.41%	325 out of 903 forecasts 35.99%
Revision per consensus (herding)	117 out of 298 forecasts 39.26%	85 out of 315 forecasts 26.98%	79 out of 290 forecasts 27.24%	281 out of 903 forecasts 31.12%
Revision against the consensus (anti-herding)	5 out of 298 forecasts 1.68%	14 out of 315 forecasts 4.44%	12 out of 290 forecasts 4.14%	31 out of 903 forecasts 3.43%
Other revisions	3 out of 298 forecasts 1.01%	7 out of 315 forecasts 2.22%	3 out of 290 forecasts 1.03%	13 out of 903 forecasts 1.44%
Subjects not exhibiting herding behavior in all rounds	16 out of 60 subjects 26.67%	20 out of 63 subjects 31.75%	20 out of 58 subjects 34.48%	56 out of 181 subjects 30.94%

"Other revisions" are ones which are in the direction of the consensus forecast, but go so far beyond it that they are further from the consensus forecast than they were before the revision.[10]

Table 4: Overview of the variance of the forecasts

	Treatment „neutral"	Treatment „positive"	Treatment „negative"	All treatments
Average standard deviation preliminary forecast	45.08	34.74	46.63	41.25
Average standard deviation final forecast	34.92	23.34	21.00	27.14

This result is further substantiated when the standard deviations of the preliminary and the final forecasts are compared (Tab. 4). The preliminary forecasts have an average standard deviation of 41.25, while the average standard deviation of the final forecast is only 27.14. The Wilcoxon signed rank test shows that the standard deviation is reduced significantly by the

---

[10] Here is an example of an "other revision": the subject's preliminary forecast is €200. The consensus forecast is €220. The subject's final forecast is then €250. This is not anti-herding, because the forecast was changed in the direction of the consensus forecast. However, the subject cannot be accused of herding behavior either, because the final forecast is further from the consensus forecast than the preliminary forecast was.

revision of the forecasts (z = 2.023; p = 0.0431). <u>Null hypothesis 3</u> thus has to be rejected too. This is a further clear indication of the occurrence of herding behavior.

In addition, it can be seen that mood really does exercise an influence on the tendency towards rational behavior. However, this influence turns out to be quite different to what was initially expected. The results of previous research showed that a neutral mood is best suited to promoting the pursuit of rational action. Positive and negative moods tend to reduce the ability and/or willingness to act rationally, it was thought. This experiment, however, reveals a completely different picture (Tab. 2). In the treatment "neutral", only 84 of the 298 preliminary forecasts (28.19%) are based on the rational approach. In the treatment "positive", the figure is slightly higher at 108 out of 315 preliminary forecasts (34.29%). In the treatment "negative", it is 100 out of 290 preliminary forecasts (34.48%). The picture is very similar for the final forecasts (Tab.2). In the treatment "neutral", only 77 of the 298 preliminary forecasts (25.84%) are based on a rational approach. In the treatment "positive", the figure is somewhat higher at 98 out of 315 preliminary forecasts (30.79%). In the treatment "negative", it is 104 out of 290 preliminary forecasts (35.86%). The Wilcoxon rank sum test shows that the difference between the treatment "neutral" and the treatment "positive" (p = 0.1669) is not significant, and neither is the difference between the treatments "positive" and "negative" (p = 0.6707). Only the treatment "neutral" differs significantly from the treatment "negative" (p = 0.0694). <u>Null hypothesis 4</u> therefore cannot be rejected. The presumption that a neutral mood is more conducive to a rational strategy than a positive or negative mood was therefore not confirmed. The opposite was closer to the truth.

It can be seen here that mood really does have an influence on the tendency towards herd behavior (Tab. 3). In the treatment "neutral", 117 out of 298 forecasts (39.26%) were corrected in the direction of the consensus forecast, while in the treatment "positive", only 85 out of 315 forecasts (26.98%) were revised in the direction of the consensus forecast. In the treatment "negative", 79 out of 290 forecasts (27.24%) were adjusted in the direction of the consensus forecast. The Wilcoxon rank sum test shows that the difference in terms of herding behavior between the treatment "neutral" and the treatment "positive" (p = 0.0333) is significant, as is the difference between the treatments "neutral" and "negative" (p = 0.0438). There is no significant difference between the treatments "negative" and the treatment "positive" (p = 0.9143). <u>Null hypothesis 5</u> thus has to be rejected. The presumption that mood does have an influence on the tendency towards herding behavior was thus confirmed. It can be seen that subjects in a neutral mood are particularly susceptible to herding behavior.

4 Summary

In this experiment, the influence of the subjects' mood on their tendency towards herding behavior in the context of share price forecasts was analyzed. The future share price is composed - in a deterministic way - of four fundamental influencing factors which the subjects are informed of before every round of the game. A random influence with an expected value of zero and with the greatest probability for the event of ±0 makes forecasting more difficult. Subjects who act rationally forecast the future share price in accordance with the four fundamental influencing factors and presume a random event of $\epsilon = \pm 0$. In this way they can maximize their expected payment. First of all the subjects make a preliminary forecast. Then they are informed of the average of all of the preliminary forecasts (consensus forecast), after which they have the opportunity to revise their preliminary forecast.

The experiment is carried out using three treatments (neutral, positive and negative) The three treatments differ in terms of the mood which is created among the subjects. The subjects' mood is influenced by film excerpts which are shown before every preliminary and every final forecast is made. Manipulation checks confirm whether the film excerpts have achieved the desired mood change.

The experiment produced the following results:

1. The subjects did not act in accordance with *homo oeconomicus*, a figurative human characterized by the ability to make rational decisions, because they did not always choose the rational strategy. Less than a third of all forecasting decisions (31.62%) followed the rational strategy, and less than a tenth of all subjects (9.94%) always made rational decisions.

2. After the consensus forecast had been announced, the subjects frequently adjusted their forecasts (in 35.99% of all cases). These adjustments were almost exclusively in the direction of the consensus forecast. A clear tendency towards herding behavior was thus observed.

3. The average standard deviation of the preliminary forecasts was significantly higher than the average standard deviation of the final forecasts. This is also a sign of herding.

4. There was hardly any sign of anti-herding. Only 3.43% of the revisions of forecasts were not in the direction of the consensus forecast, and were thus in the opposite direction.

5. The rational strategy was followed significantly more often in the treatments "positive" (34.29%) and negative (34.48%) than in the treatment "neutral" at 28.19%.

6. Adjustments of the forecasts in the direction of the consensus forecast (herding behavior) occurred in the treatments "positive" (26.98%) and "negative" (27.24%) significantly less often than in the treatment "neutral" (39.26%).

In summary, it can be stated that herding can be observed very frequently. The mood of the subjects has an influence on the occurrence of herding. The tendency towards herding behavior is stronger in a neutral mood than in a positive or negative mood. In addition, less attention is paid to the rational strategy in a neutral mood than in a positive or negative mood.

References

Allwood, C. M. / Granhag, P. A. / Jonsson, A. C. (2002): Does Mood Influence the Realism of Confidence Judgements?, in: Scandinavian Journal of Psychology, Bd. 43, H. 3, S. 253-260.

Anderson, L. R. / Holt, C. (1997): Information Cascades in the Laboratory, in: The American Economic Review, Bd. 87, H. 5, S. 847-862.

Andrade, E. B. / Odean, T. / Lin, S. (2016): Bubbling with Excitement: An Experiment, in: Review of Finance, Bd. 20, H. 2, S. 1-20.

Apergis, N. / Gabrielsen, A. / Smales, L. A. (2016): (Unusual) Weather and Stock Returns—I Am Not in the Mood for Mood: Further Evidence from International Markets, in: Financial Markets and Portfolio Management, Bd. 30, H. 1, S. 63-94.

Ashiya, M. / Doi, T. (2001): Herd behavior of Japanese economists, in: Journal of Economic Behavior and Organization, Bd. 46, N. 3, S. 343-346.

Baddeley, M. / Curtis, A. / Wood, R. (2004): An Introduction to Prior Information Derived from Probabilistic Judgments: Elicitation of Knowledge, Cognitive Bias and Herding (Special Publication), Bd. 239, Geological Society, London, S. 15-27.

Baddeley, M. (2010): Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses, in: Philosophical Transactions, The Royal Society B, Bd. 365, H. 1538, S. 281-290.

Baker, M. / Wurgler, J. (2007): Investor Sentiment in the Share Market, in: Journal of Economic Perspectives, Bd. 21, H. 2, S. 129-151.

Banerjee, A. (2010): A simple model of herd behavior, in: Quarterly Journal of Economics, Bd. 107, H. 3, S. 797-817.

Baumeister, R. F. / Vohs, K. D. / Dewall, C. N. / Zhang, L. (2007): How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, rather than Direct Causation, Personality, in: Social Psychology Review, Bd. 11, H. 2, S. 167-203.

Bassi, A. / Colacito, R. / Fulghieri, P. (2013): O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions, in: Review of Financial Studies, Bd. 26, H. 7, S. 1824-1852.

Becker, O. / Leitner, J. / Leopold-Wildburger, U. (2009): Expectation formation and regime switches, in: Experimental Economics, Bd. 12, S. 350-364.

Bedke, N. / Bizer, K. / Spiwoks, M. (2009): Gregarious Analysts – Experimental Evidence for Reputational Herding, in: Journal of Money Investment and Banking, Bd. 12, S. 26-36.

Bernhardt, D. / Campello, M. / Kutsoati, E. (2006), Who herds? in: Journal of Financial Economics, Bd. 80, H. 3, S. 657-675.

Bhalla, M. (2012): Social Learning Among Rational Analysts, in: The Journal of Behavioral Finance, Bd. 13, S. 164-173.

Bikhchandani, S. / Hirshleifer, D. / Welch, I. (1992): A theory of fads, fashion, custom, and cultural change as informational cascades, in: Journal of Political Economy, Bd. 100, H. 5, S. 992-1026.

Bikhchandani, S. / Sharma, S. (2000), Herd behavior in financial markets, IMF Staff Papers, Bd. 47, H. 3, pp. 279-310.

Black, F. (1986): Noise, in: Journal of Finance, Bd. 41, H. 3, S. 529-543.

Bless, H. / Schwarz, N. / Kemmelmeier, M. (1996): Mood and Stereotyping: The Impact of Moods on the use of General Knowledge Structures, in: European Review of Social Psychology, Bd. 7, H. 1, S. 63-93.

Breaban, A. / Noussair, C. N. (2018): Emotional State and Market Behavior, in: Review of Finance, Bd. 22, H. 1, S. 279–309.

Cao, M. / Wei, J. (2005): Stock Market Returns: A Note on Temperature Anomaly, in: Journal of Banking and Finance, Bd. 29, H. 6, S. 1559–1573.

Celen, B. / **Kariv, S. (2004)**: Distinguishing Informational Cascades from Herd Behavior in the Laboratory, in: American Economic Review, Bd. 94, H. 3, S. 484-498.

Chang, S. C. / Chen, S. S. / Chou, R. K. / Lin, Y. H. (2008): Weather and Intraday Patterns in Stock Returns and Trading Activity, in: Journal of Banking and Finance, Bd. 32, H. 9, S. 1754-1766.

Chang, T. / Nieh, C. C. / Yang, M. J. / Yang, T. Y. (2006): Are Stock Market Returns Related to the Weather Effects? Empirical evidence from Taiwan, in: Physica A: Statistical Mechanics and its Applications, Bd. 364, S. 343–354.

Clement, M. B. / Tse, S. Y. (2005), Financial analyst characteristics and herding behavior in forecasting, in: The Journal of Finance, Bd. 60, N. 1, S. 307-341.

Colasante, A. / Marini, M. / Russo, A. (2017): Incidental Emotions and Risk-Taking: An Experimental Analysis, MPRA Paper No. 82767.

Cote, J. M. / Sanders, D. L. (1997): Herding Behavior: Explanations and Implications, in: Behavioral Research in Accounting, Bd. 9, S. 20-45.

Daniel, K. / Hirshleifer, D. / Teoh, S. H. (2002): Investor Psychology in Capital Markets: Evidence and Policy Implications, in: Journal of Monetary Economics, Bd. 49, H. 1, S. 139-209.

Devenow, A. / Welch, I. (1996): Rational herding in financial economics, in: European Economic Review, Bd. 40, H. 3-5, S. 603-615.

Dowling, M. / Lucey, B. M. (2005): Weather, Biorhythms, Beliefs and Stock Returns – some preliminary Irish evidence, in: International Review of Financial Analysis, Bd. 14, H. 3, S. 337–355.

Duxbury, D. (2015): Behavioral finance: insights from experiments II: biases, moods and emotions, in: Review of Behavioral Finance, Bd. 7, H. 2, S. 151–175.

Elster, J. (1996): Rationality and the Emotions, in: The Economic Journal, Bd. 106, H 438, S. 1386-1397.

Elster, J. (1998): Emotions and Economic Theory, in: Journal of Economic Literature, Bd. 36, H. 1, S. 47–74.

Filiz, I. (2017): Overconfidence: the Influence of Positive and Negative Affect, Wolfsburg Working Papers, No. 17–02, Wolfsburg.

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

Floros, C. (2011): On the Relationship between Weather and Stock Market Returns, in: Studies in Economics and Finance, Bd. 28, H. 1, S. 5-13.

Galariotis, E. C. / Rong, W. / Spyrou, S. I. (2015): Herding on fundamental information: A comparative study, in: Journal of Banking & Finance, Bd. 50, S. 589-598.

Gavriilidis, K. / Kallinterakis, V. / Tsalavoutas, I. (2016): Investor mood, herding and the Ramadan effect, in: Journal of Economic Behavior & Organization, Bd. 132, S. 23-38.

George, J. M. / Dane, E. (2016): Affect, emotion, and decision making, in: Organizational Behavior and Human Decision Processes, Bd. 136, S. 47-55.

Grable, J. E. / Roszkowski, M. J. (2008): The influence of mood on the willingness to take financial risks, in: Journal of Risk Research, Bd. 11, H. 7, S. 905-923.

Graham, J. R. (1999): Herding among investment newsletters: theory and evidence, in: Journal of Finance, Bd. 54, S. 237-268.

Hirshleifer, D. (2001): Investor Psychology and Asset Pricing, in: The Journal of Finance, Bd. 56, H. 4, S. 1533-1597.

Hirshleifer, D. / Shumway., T. (2003): Good day sunshine: Share returns and the weather, in: Journal of Finance, Bd. 58, H. 3, 1009-1032.

Hong, H. / Kubik, J. D. / Solomon, A. (2000): Security analysts' career concerns and herding of earnings forecasts, in: RAND Journal of Economics, Bd. 31, H. 1, S. 121-144.

Huck, S. / Oechssler, J. (2000): Informational cascades in the laboratory: Do they occur for the right reasons? In: Journal of Economic Psychology, Bd. 21, S. 661-671.

Hung, A. H. / Plott, C. R. (2001): Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions, in: The American Economic Review, Bd. 91, H. 5, S. 1508-1520.

Ifcher, J. / Zarghamee, H. (2014): Affect and Overconfidence: A Laboratory investigation, in: Journal of Neuroscience, Psychology, and Economics, Bd. 7, H. 3, S. 125-150.

Isen, A. (2000): Positive Affect and Decision Making, in: Handbook of Emotion, Lewis, M. / Haviland-Jones, J., New York.

Johnson, E. J. / Tversky, A. (1983): Affect, Generalization, and the Perception of Risk, in: Journal of Personality and Social Psychology, Bd. 45, H. 1, S. 20–31.

Kamstra, M. J. / Kramer, L. A. / Levi, M. D. (2003): Winter Blues: A SAD share market cycle, in: American Economic Review, Bd. 93, H. 1, S. 324-343.

Kang, S. H. / Jiang, Z. / Lee, Y. / Yoon, S.-M. (2010): Weather Effects on the Returns and Volatility of the Shanghai Stock Market, in: Physica A: Statistical Mechanics and its Applications, Bd. 389, H. 1, S. 91-99.

Kaplanski, G. / Levy, H. / Veld, C. / Veld-Merkoulova, Y. (2015): Do Happy People make optimistic Investors? In: Journal of Financial and quantitative Analysis, Bd. 50, H. 1-2, S. 145-168

Kaustia, M. / Rantapuska, E. (2016): Does Mood Affect Trading Behavior? In: Journal of Financial Markets, BD. 29, S. 1-26.

Keef, S. P. / Roush, M. L. (2007): Daily Weather Effects on the Returns of Australian Stock Indices, in: Applied Financial Economics, Bd. 17, H. 3, S. 173–184.

Keynes, J. M. (1936): The General Theory of Employment. Macmillan, Interest and Money. London.

Kim, C. / Pantzalis, C. (2003): Global/industrial diversification and analyst herding, in: Financial Analysts Journal, Bd. 59, H. 2, pp. 69-79.

Kim, J. H. (2017): Stock Returns and Investors' Mood: Good Day Sunshine or Spurious Correlation? in: International Review of Financial Analysis, Bd. 52, S. 94-103.

Kirchsteiger, G. / Rigotii, L. / Rustichine, A. (2006): Your Morals Might Be your Moods, in: Journal of Economic Behavior and Organization, Bd. 59, H. 2, S. 155-172.

Krämer, W. / Runde, R. (1997): Stocks and the Weather: An Exercise in Data Mining or Yet Another Capital Market Anomaly? In: Empirical Economics, Bd. 22, H. 4, S. 637-641.

Kraemer, C. / Nöth, M. / Weber, M. (2006): Information aggregation with costly information and random ordering, in: Journal of Economic Behavior and Organization, Bd. 59, S. 423-432.

Krishnan, M. / Lim, S. C. / Zhou, P. (2005): Who Herds? Who Doesn`t? Working Paper.

Krivelyova, A. / Robotti, C. (2003): Playing the Field: Geomagnetic Storms and International Stock Markets, Working Paper Federal Reserve Bank of Atlanta, 2003-5.

Kübler, D. / Weizsäcker, G. (2004): Limited depth of reasoning and failure of cascade formation in the laboratory, in: The Review of Economic Studies, Bd. 71, H. 2, S. 425-441.

Kuhnen, C. M. / Knutson, B. (2011): The Influence of Affect on Beliefs, Preferences, and Financial Decisions, in: Journal of Financial and Quantitative Analysis, Bd. 46, H. 3, S. 605-626.

Lahav, Y. / Meer, S. (2012): The Effect of Induced Mood on Prices in Asset Markets - Experimental Evidence, Working Paper, Ben-Gurion University of the Negev.

Lamont, O. A. (2002): Macroeconomic forecasts and microeconomic forecasters, in: Journal of Economic Behavior & Organization. Bd. 48, S. 265–280.

Lane, T. (2017): How does happiness relate to economic behavior? A review of the literature Macroeconomic, in: Journal of Behavioral and Experimental Economics. Bd. 68, S. 62–78.

Laux, C. M. / Probst, D. A. (2004): One signal, two opinions: strategic heterogeneity of analysts' forecasts, in: Journal of Economic Behavior & Organization, Bd. 55, S. 45–66.

Le Bon, G. (1896): The Crowd: A Study of the Popular Mind, T. Fisher Unwin, London.

Lee, C. J. / Aandrade, E. B. (2014): Fear, Excitement, and Financial Risk-Taking, in: Cognition and Emotion, Bd. 29, H. 1, S. 178-187.

Lerner, J. S. / Li Y. / Valdesolo, P. / Kassam, K. S. (2015): Emotion and Decision Making, in: Annual Review of Psychology, Bd. 66, H. 1, S. 799–823.

Loewenstein, G. (2000): Emotions in Economic Theory and Economic Behavior, in: American Economic Review, Bd. 90, H. 2, S. 426-432.

Loewenstein, G. / Weber, E. / Hsee, C. / Welch, N. (2001): Risk as Feelings, in: Psychological Bulletin, Bd. 127, H. 2, S. 267–286.

Lu, J. / Chou, R. K. (2012): Does the Weather have Impacts on Returns and Trading Activities in Order-Driven Stock Markets? Evidence from China, in: Journal of Empirical Finance, Bd. 19, H. 1, S. 79–93.

Lucey, B. M. / Dowling, M. (2005): The Role of Feelings in Investor Decision-Making, in: Journal of Economic Surveys, Bd. 19, H. 2, S. 211–237.

Mackay, C. (1841): Extraordinary Popular Delusions and the Madness of Crowds. Richard Bentley, London.

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

Mirza, H. H. / Asghar, A. / Jam E Kausar, M. / Mushtaq, N. (2012): Stock Market Returns and Weather Anomaly: Evidence from an Emerging Economy, in: Journal of Economics and Behavioral Studies, Bd. 4, H. 5, S. 239-244.

Naujoks, M. / Aretz, K. / Kerl, A. / Walter, A. (2009): Do German security analysts' herd?, in: Financial Markets and Portfolio Management, Bd. 23, H. 1, S. 3–29.

Nofsinger, J. R. (2005): Social Mood and Financial Economics, in: The Journal of Behavioral Finance, Bd. 6, H. 3, S. 144-160.

Nöth, M. / Weber, M. (2003): Information Aggregation with Random Ordering: Cascades and Overconfidence, in: The Economic Journal, Bd. 113, S. 166-189.

Olsen, L. (1996): Implications of herding behavior for earnings estimation, risk assessment, and stock returns, in: Financial Analysts Journal, Bd. 52, N. 4, S. 37-41.

Orlean, A. (1995): Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion, in: Journal of Economic Behavior and Organization, Bd. 28, S. 257-274.

Oswald, A. / Proto, E. / Sgroi, D. (2015): Happiness and Productivity, in: Journal of Labor Economics, Bd. 33, H. 4, S. 789-822.

Pardo A. / Enric, V. (2002): Spanish Stock Returns: Where is the Weather Effect? In: European Financial Management, Bd. 9, H. 1, S. 117-126.

Pham, M. T. (2007): Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence, in: Review of General Psychology, Bd. 11, H. 2, S. 155-178.

Pierdzioch, C. / Rülke, J.-C. (2012): Forecasting stock prices: do forecasters herd? In: Economics Letters, Bd. 116, H. 3, S. 326-329.

Pierdzioch, C. / Rülke, J.-C. / Stadtmann, G. (2013): Forecasting metal prices: do forecasters herd? In: Journal of Banking Finance, Bd. 37, H.1, S. 150-158.

Raafat, R. M. / Chater, N. / Frith, C. (2009): Herding in humans, in: Trends in Cognitive Sciences, Bd. 13, H. 10, S. 420-428.

Roider, A. / Voskort, A. (2016): Reputational Herding in Financial Markets: A Laboratory Experiment, in: Journal of Behavioral Finance, Bd. 17, H. 3, S. 244-266.

Rottenberg, J. / Ray, R. D. / Gross, J. J. (2007): Emotion Elicitation Using Films, in: Coan, J. A., Allen, J. J. B., in: The Handbook of Emotion Elicitation and Assessment, London, S. 9-28.

Russell, G. W. (2004): Sport riots: A social-psychological review. Aggression and Violent Behav. 9, S. 353-378.

Sasaki, S. (2005): Signal Qualities, Order of Decisions, and Informational Cascades: Experimental Evidence in: Economics Bulletin, Bd. 3, H. 34, S. 1-11.

Saunders, E. M. (1993): Stock Prices and Wall Street weather, in: American Economic Review, Bd. 83, H. 5, S. 1337–1345.

Scharfstein, D. S. / Stein, J. C. (1990): Herd behavior and investment, in: The American Economic Review, Bd. 80, H. 3, S. 465-479.

Sschwarz, N. (2000): Emotion, Cognition, and Decision Making, in: Cognition & Emotion, Bd. 14, H. 4, S. 433–440.

Sgroi, D. (2003): The Right Choice at the Right Time: A Herding Experiment in Endogenous Time, in: Experimental Economics, Bd. 6, S. 159-180.

Shiv, B. / Loewenstein, G. / Bechara, A. / Damasio, H. / Damasio, A. R. (2005): Investment Behavior and the Negative Side of Emotion, in: Psychological Science, Bd. 16, H. 6, S. 435-439.

Shleifer, A. / Summers, L. (1990): The noise trader approach to finance, in: The Journal of Economic Perspectives, Bd. 4, H. 2, S. 19-33.

Shu, H.-C. / Hung, M. W. (2009): Effect of Wind on Stock Market Returns: Evidence from European Markets, in: Applied Financial Economics, Bd. 19, H. 11, S. 893–904.

Simonsohn, U. / Ariely, D. (2008): When rational sellers face nonrational buyers: evidence from herding on eBay, in: Management Science, Bd. 54, H. 9, S. 1624-1637.

Smith, A. (1759): The theory of moral sentiments, A. Millar.

Spiwoks, M. / Bizer, K. / Hein, O. (2008a): Informational Cascades: A Mirage?, in: Journal of Economic Behavior & Organization, Bd. 67, H. 1, 2008, S. 193-199.

Spiwoks, M. / Bizer, K. / Hein, O. (2008b): Anchoring near the Lighthouse: Bond Market Analysts' Behavior Co-ordination by External Signal, in: European Journal of Economics, Finance and Administrative Science, Bd. 13, S. 169-191.

Spyrou, D. (2013): Herding in financial markets: a review of the literature, in: Review of Behavioral Finance, Bd. 5, H. 2, S. 175-194.

Stickel, S. E. (1990): Predicting Individual Analyst Earnings Forecasts, in: Journal of Accounting Research, Bd. 28, H. 2, S. 409-417.

Sumpter, D. J. T. / Zabzina, N. / Nicolis, S. C. (2012): Six Predictions about the Decision Making of Animal and Human Groups, in: Managerial and Decision Economics, Bd. 33, S. 295-309.

Symeonidis, L. / Daskalakis, G. / Markellos, R. N. (2010): Does the Weather Affect Stock Market Volatility? In: Finance Research Letters, Bd. 7, H. 4, S. 214-223.

Tufan, E. / Hamarat, B. (2004): Do Cloudy Days Affect Stock Exchange Returns: Evidence from the Istanbul Stock Exchange, in: Journal of Naval Science and Engineering, Bd. 2, H. 1, S. 117–26.

Vohs, K. / Baumeister, R. / Loewenstein, G. (2007): Do emotions help or hurt decision making? A hedgefoxian perspective. New York.

Waller, J. (2009): The art of medicine A forgotten plague: making sense of dancing mania, in: Lancet 373, S. 624-625.

Watson, D. / Vaidya, J. (2003): Mood measurement: Current status and future directions, in: Handbook of Psychology, Bd. 2: Research methods in psychology, New York, John Wiley, S. 351-375.

Westermann, R. / Spies, K. / Stahl, G. / Hesse, A. (1996): Relative Effectiveness and Validity of Mood Induction Procedures: A Meta-analysis, in: European Journal of Social Psychology, Bd. 26, H. 4, S. 557-580.

Zitzewitz, E. (2001): Measuring herding and exaggeration by equity analysts and other opinion sellers, Stanford GSB Working Paper No. 1802, Stanford.

Appendix: instructions and test questions

Instructions

_____   _____
Player number   Computer number

The Game

In this game you are requested to make forecasts for the future trend of a share price. This share price in euros at the point in time t (K_t) is always determined by four influencing factors (the fundamental data A_t, B_t, C_t and D_t) and a random influence ($\mathcal{E}_t$). The fundamental data is provided before each forecast is made. The subjects are also aware of the specific influence the fundamental data has on the share price.

$$K_t = 2 \cdot A_t + 3 \cdot B_t - 1 \cdot C_t + 4 \cdot D_t + \mathcal{E}_t$$

The value of the random influence ($\mathcal{E}_t$) moves within a range of €-40 and €+40. The probability of the events €-40, €-30, €-20, €+20, €+30 and €+40 is 10% each. The probability of the events €-10 and €+10 is 12%. The event €±0 has a probability of occurrence of 16% (Figure 1).

Fig. 1: Distribution of the probability of the random influence $\mathcal{E}_t$ on the formation of the share price in euros
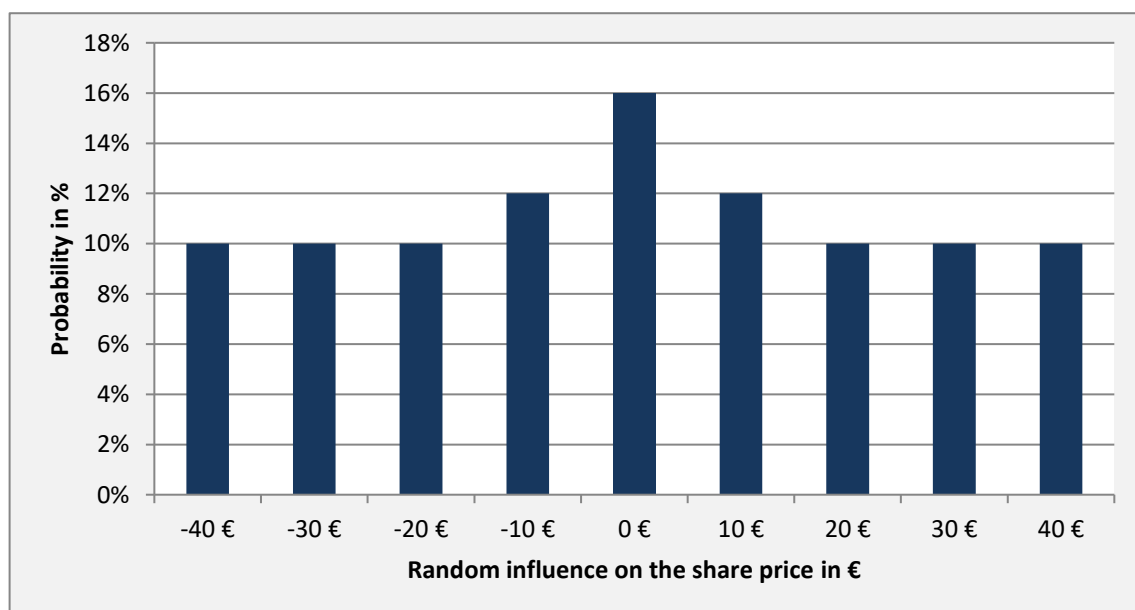
The random influence ϵ_t for each period is determined by a lottery. On the next page you can see the random influences of the last 10 periods (Fig. 2).

Fig. 2: Random influences ϵ_t of the last 10 periods
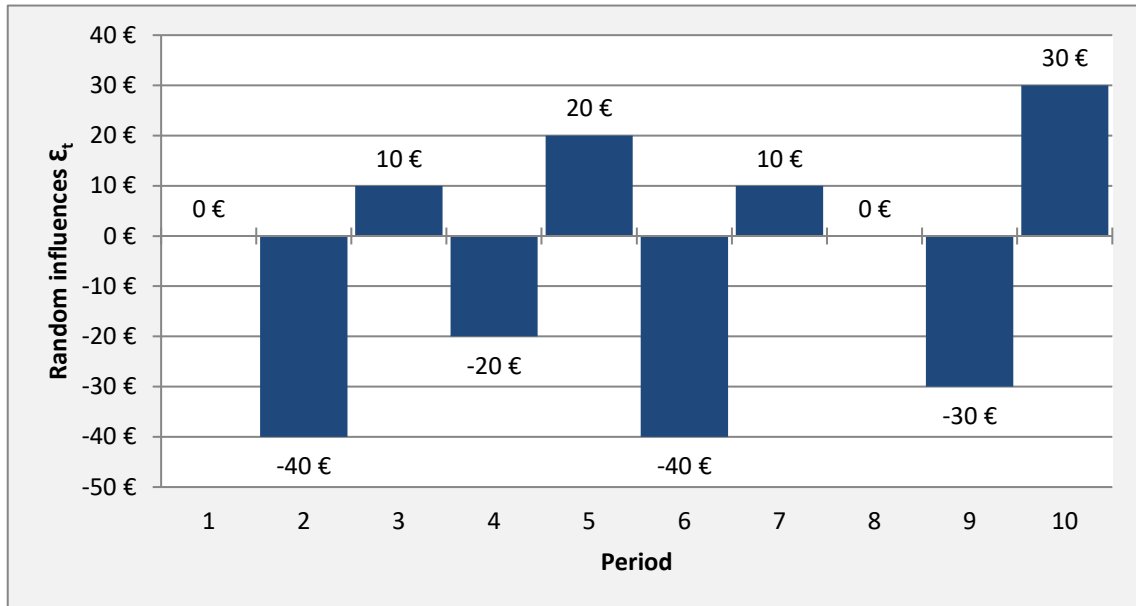
As an example, here is how the share prices were determined for Period 6.

The fundamental data and the random influence are as follows:

A_6 = €31; B_6 = €26; C_6 = €25; D_6 = €31; ϵ_6 = €-40

Hence:
$K_6 = 2 \cdot €31 + 3 \cdot €26 - 1 \cdot €25 + 4 \cdot €31 + €\text{-}40 = €199$

In Figure 3 you can see the price movements of the share over the last ten periods.

Fig. 3: Price movements of the share over the last ten periods.
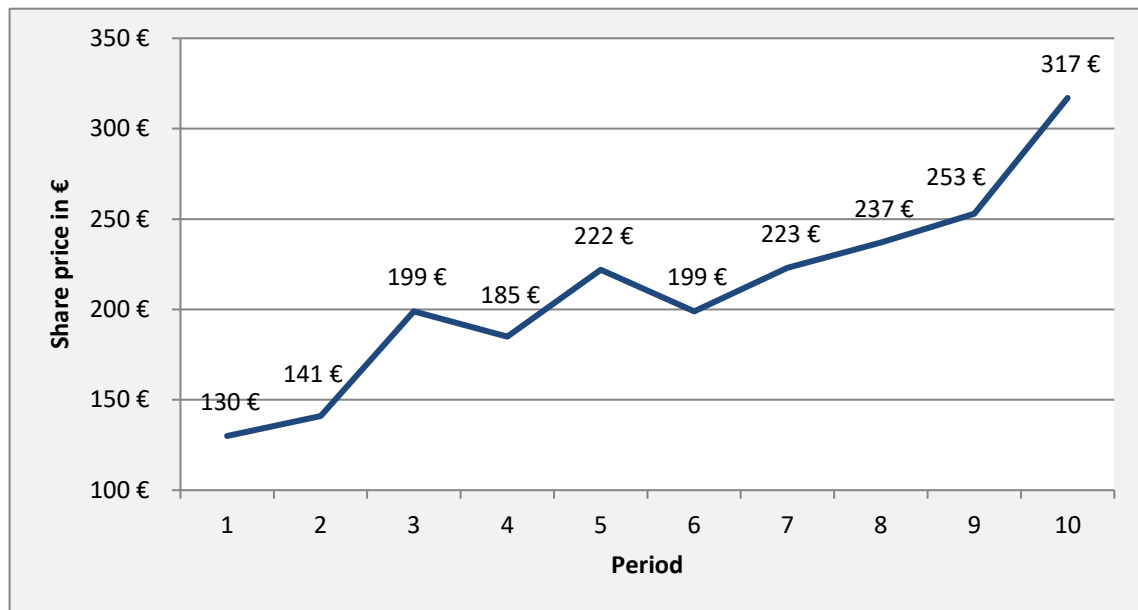
Share price in euros / period

Procedure:

After reading the instructions and answering the test questions you will see a film excerpt. Following this you will receive the figures of the fundamental data for the current period. You will be asked to make a forecast for the share price. Then you will see another film excerpt. Then you will receive the information about the consensus forecast. The consensus forecast is the arithmetical average of all preliminary forecasts made by the participants. Now you have the opportunity to revise your forecast. A total of five rounds are played. Before each round you see the trajectory of the share price and the random influences of the last rounds of the game.

Payment:

You will receive a show-up fee of €5 for taking part in the experiment. For every successful share price forecast you receive €10. A forecast is considered successful and is rewarded accordingly when it does not diverge by more than €10 from the actual share price. In total you can earn up to €55. Payment is made at the end of the experiment.

Information:

- Please remain quiet during the experiment
- Do not look at your neighbors' screen
- Apart from a pocket calculator, no aids are permitted (smartphones, smart watches etc.)
- Please note the respective time limits given on the upper right of the screen.

Test questions

<u>**Multiple-choice test questions:**</u>

1. What is your task in this game?
   a. Solving mathematical problems.
   b. Making forecasts for the future trend of a share price. (correct)
   c. Making diversification decisions.

2. What is the probability of occurrence of the random influences €-10 and €+10?
   a. 16%
   b. 10%
   c. 12% (correct)

3. What does the payment depend on?
   d. On the movement of the share price.
   e. On the success of the forecast. (correct)
   f. On the level of the DAX.

4. The current share price is €232 and you forecast €241. How much is your payout?
   a. €5
   b. €0
   c. €10 (correct)

Chapter 6

THE ACCURACY OF INTEREST RATE FORECASTS IN THE ASIA-PACIFIC REGION: OPPORTUNITIES FOR PORTFOLIO MANAGEMENT

with Ibrahim Filiz, Markus Spiwoks and Kilian Bizer
contribution Thomas Nahmer: 35%

Abstract

We analyzed interest rate forecasts from Australia, China, Hong Kong, India, Indonesia, Malaysia, New Zealand, Singapore, South Korea, Taiwan and Thailand. We assessed 532 forecast time series with a total of 85,264 individual interest rate forecasts. To do so, we carried out a comparison to naïve forecasts and investigated the forecast time series for topically-orientated trend adjustments. In addition, we deployed the sign accuracy test and the unbiasedness test. The results are very sobering in part: 95.9% of all forecast time series are characterized by the phenomenon of topically-orientated trend adjustments, and 99.4% of all forecast time series proved to be biased. Only a small proportion of the forecast time series (3.6%) reflected the future interest rate trend significantly more precisely than a naïve forecast. However, at the same time some of the results of the study are surprisingly positive. The sign accuracy test revealed that 48.3% of all forecast time series predict the interest rate trend significantly better than a random walk forecast.

Keywords

Interest rate forecasts, survey forecasts, forecast accuracy, portfolio management, topically-orientated trend adjustment behavior.

JEL Classification

E44, E47, G11, G12, G15, G17, G21, F37

1 Introduction

Future interest rate trends are of key significance for almost all investment decisions on the capital markets. If the general level of interest rates in an economy rises, the prices of most bonds will fall. Only securities with a short residual term to maturity and floating-rate bonds remain largely unaffected by such developments. The longer the term to maturity and the lower the coupon, the greater the fall in bond prices. If the general level of interest rates falls, the opposite effect occurs, and the prices of most bonds rise.

Interest rate trends are, however, also of great importance for investments in the stock market. The fair value of a share is the total of all discounted future profits that this share provides. If one wishes to take a critical look at the current market price of a stock, it is wise not to compare its current price with the current fair value, but rather to determine the *future* fair value of the stock. If the current value is significantly below the fair value which it will have at the end of the investment horizon, it is an attractive investment. However, in order to establish the future fair value of a stock, one has to forecast the predominant future interest rate level, because this simultaneously represents the future discount rate in the determination of fair value.

In the case of international portfolios, exchange rate movements also have to be taken into account. The interest rate parity theory shows that interest rate trends at home and abroad have great significance for exchange rates.

As a rule, financial market analyzes normally begin with a forecast of interest rate trends, because bond and share prices – and ultimately also exchange rates – are significantly influenced by interest rates. It is therefore not surprising that the accuracy of interest rate forecasts have been of great interests to academics and business professionals for a long time now. In the past 40 years, around 50 studies have already been published on the accuracy of survey-based interest rate forecasts (see the comprehensive synoptic overview in Table 20 in the appendix). Some trends have emerged in these studies:

Only a few studies considered the interest rate forecasts they analyzed to be largely reliable.[1] These were largely forecasts on the base rates of central banks or forecasts of short-term market rates such as the three-month money market rate. However, for portfolio management, interest rate forecasts for bonds with maturities of at least a year are primarily of interest, because active portfolio management strategies can be realized much more easily in this segment. There are also some studies with mixed findings.[2] Here again, it is mostly forecasts

---

[1] See, for example, Throop (1981), Tabak and Feitosa (2008), Baghestani and Marchon (2012), Knüppel and Schultefrankenfeld (2013) and Pierdzioch (2015).

[2] See, for example, Dua (1988), Zarnowitz and Braun (1992), Cho (1996), Gosnell and Kolb (1997), Greer (2003), Scheier and Spiwoks (2006), Goodhart and Lim (2008), Spiwoks, Bedke and Hein (2008), Chun (2009), Spiwoks, Bedke and Hein (2010), Jongen, Verschoor and Wolff (2011), Kunze, Kramer and Rudschuk (2013), Kunze and Gruppe (2014), Baghestani and Danila (2014), Beechay and Österholm (2014), Oliver and Pasaogullari (2015), and Miah, Khalifa and Hammoudeh (2016).

for short-term interest rates which come off well, while more than half of the studies take a very critical view of the quality of the interest rate forecasts which they examined.[3]

Until now, US interest rate forecasts have been the main focus of research, although European interest rate forecasts – particularly British and German ones – have also been frequently examined. In the Asia-Pacific region there have been several studies focusing on Japan,[4] but otherwise there has only been little published research dealing with interest rate forecasts for the Asia-Pacific region. Goodhart and Lim (2008) looked at interest rate forecasts in New Zealand, while Baghestani, Arzaghi and Kaya (2015) analyzed their Australian counterparts. Jongen, Verschoor and Wolff (2011) investigated forecasts of interest rate trends in Hong Kong, Indonesia, Malaysia, New Zealand, Singapore and Taiwan for the period 1995-2009. However, this study limited itself to assessment criteria which do not provide a comprehensive impression of the accuracy of the forecasts. Instead it compared survey forecasts with random walk forecasts or implicit forward rate forecasts. Miah, Khalifa and Hammoudeh (2016) analyzed interest rate forecasts from China, Hong Kong, India, Indonesia, the Philippines, Singapore, South Korea, Taiwan and Thailand, among others. They examined the period 2001-2012 and applied the efficiency test and the unbiasedness test. As a data basis they used the survey forecasts of Fx4casts.com. For us, this is definitely the most interesting study among those which have previously been carried out. However, we refer to a different data basis (Asia Pacific Consensus Forecasts), and in addition to interest rate forecasts from China, Hong Kong, India, Indonesia, Singapore, South Korea, Taiwan and Thailand, we also look at forecasts from Malaysia, Australia and New Zealand. We also take a longer period of time into account in our analysis (1990-2016).[5] In addition, we use a far more comprehensive set of tools for the evaluation of the quality of forecasts: (1) Comparison to a naïve forecast with the aid of the Diebold-Mariano test, (2) examination of the forecast direction with the help of the sign accuracy test, (3) examination for systematic forecast errors with the aid of the unbiasedness test, and (4) test for the presence of possible topically-oriented trend adjustments with aid of the TOTA coefficient.

Unlike many of the previous studies we not only examine the time series of the consensus forecasts, but also the forecast time series of the banks, investment companies, associations, consulting firms and industrial companies which participated in the survey. It cannot be ruled out that individual institutions might succeed in making forecasts which are more reliable than the consensus forecasts. Limiting the analysis to consensus forecasts might therefore mislead us. We evaluated a total of 532 time series with 85,264 interest rate forecasts. In this respect,

---

[3] See, for example, Friedman (1980), Belongia (1987), Simon (1989), Hafer and Hein (1989), Francis (1991), Hafer, Hein and MacDonald (1992), Domian (1992), Ilmanen (1996), Kolb and Stekler (1996), Baghestani, Jung and Zuchegno (2000), Albrecht (2000), Spiwoks (2003), Brooks and Gray (2004), Benke (2004), Mose (2005), Baghestani (2005), Benke (2006), Spiwoks and Hein (2007), Mitchell and Pearce (2007), Spiwoks, Bedke and Hein (2009), Gubaydullina, Hein and Spiwoks (2011), Schwarzbach, Kunze, Rudschuck and Windels (2012), Chortareas, Jitmaneeroj and Wood (2012), Butter and Jansen (2013), Spiwoks, Gubaydullina and Hein (2015), and Kunze, Wegener, Bizer and Spiwoks (2017).

[4] See, for example, Gosnell and Kolb (1997), Spiwoks and Hein (2007), Gubaydullina, Hein and Spiwoks (2011), Jongen, Verschoor and Wolff (2011), Butter and Jansen (2013), Spiwoks, Gubaydullina and Hein (2015), and Baghestani, Arzaghi and Kaya (2015).

[5] The Australian interest rate forecasts start in 1990. The other time series only begin later.

the study can be viewed as the most comprehensive analysis by far of interest rate forecasts in the Asia-Pacific region.

Some surprising results were revealed in the process, which certainly opens up opportunities for active portfolio management strategies. For example, 61.5% of the forecast time series on the interest rates of Indian state bonds with ten years remaining to maturity (forecast horizon: 13 months) predict the future interest rate trend (rising or falling) significantly better than a random walk forecast. With forecasting results of this kind, it should be possible to systematically obtain excess returns.

The study is divided into five sections. In chapter 2 the data basis is described in detail. In chapter 3 the methods used are presented. The results of the study are shown in chapter 4. In the final chapter a summary of the study is provided.

2 Data Basis

Bates and Granger (1969) were the first to raise the question of whether better results could be achieved via a suitable combination of several forecasts than by means of choosing the (presumably) best forecast. The idea behind this is that many forecasts contain useful elements of information which are not found in other forecasts and which can be brought together in a consensus forecast (see, for example, Thiele 1993). This idea triggered a lively debate on the possibilities and limitations of suitable combinations of forecasts, which culminated in 1989 in a special edition of the Journal of Forecasting and the International Journal of Forecasting. As a result of this debate, the company Consensus Economics created the specialist journal Consensus Forecasts. Since October 1989 it has been published on a monthly basis. Every month, Consensus Economics interviews more than 700 leading academics from the fields of economics and business for their forecasts in relation to various economic indicators for over 85 countries. Alongside the forecasts of these experts, Consensus Economics also publishes a consensus mean, which is the arithmetical average of the experts' forecasts.

The interest rate forecasts for Australia, China, Hong Kong, India, Indonesia, Malaysia, New Zealand, Singapore, South Korea, Taiwan and Thailand which are analyzed here come from the regularly published journal Asia Pacific Consensus Forecasts. We examined the forecasts which were published there in the period from January 1990 to December 2015. The forecasts relate to the period from April 1990 to the end of December 2016. The data for Australia in the initial years is from the journal Consensus Forecasts. For the period after the establishment of the journal Asia Pacific Consensus Forecasts in 1995, the Australian interest rate forecasts are taken from that periodical. We evaluated a total of 532 time series with 85,264 interest rate forecasts. There is a detailed overview in Table 1.

Asia Pacific Consensus Forecasts differentiates between two forecast horizons: In the journal, the forecasts are occasionally described as three-month forecasts and twelve-month forecasts. In reality, however the forecast horizons are of four and thirteen months. This can be seen in the following example: In the edition of January 2015, which was available in around mid-January, the forecasts for the end of April 2015 and the end of January 2016 were

published. The forecasts themselves are handed in by the participating institutions at the beginning of January. The actual period of time from the beginning of January 2015 to the end of April 2015 is four months, while the period of time from the beginning of January 2015 to the end of January 2016 is thirteen months.

Table 1: Data Used from the Journal Asia Pacific Consensus Forecasts

Country	Subject of Forecast	Number of Time Series Analyzed	Number of Forecasts Analyzed	Results in the Table
Australia	10 Year Government Bond Yield	42	7,871	3
	Three Month Interest Rates	42	8,115	4
China	One-Year Base Lending Rate	30	3,507	5
Hong Kong	Prime Lending Rate	30	5,159	6
	Three Month Interest Rates	38	6,077	7
India	10 Year Government Bond Yield	26	3,809	8
	Three Month Interest Rates	24	4,196	9
Indonesia	10 Year Government Bond Yield	24	3,595	10
Malaysia	Base Lending Rate	30	4,374	11
	Three Month Interest Rates	36	5,842	12
New Zealand	10 Year Government Bond Yield	36	6,566	13
	Three Month Interest Rates	36	6,552	14
Singapore	Prime Lending Rate	30	3,876	15
	Three Month Interest Rates	38	5,906	16
South Korea	Three Year Interest Rates	28	3,194	17
Taiwan	10 Year Government Bond Yield	16	2,103	18
Thailand	Three Month Interest Rates	26	4,522	19
∑		532	85,264	

We analyzed all of the forecast time series which have at least 80 items of data. We did not take time series with less than 80 observations into consideration. Under certain circumstances, time series which are too short or contain too large gaps can lead to inconclusive results in the procedures used to measure the quality of forecasts.

3 Methods

The following statistical tools were used to measure the quality of forecasts: Comparison to a naïve forecast with the aid of the Diebold-Mariano test (3.1), examination of the forecast direction with the help of the sign accuracy test (3.2), the test for the unbiasedness of the forecasts (3.3) and the test for the presence of topically-orientated trend adjustments with the help of the TOTA coefficient (3.4).

3.1 Comparison to a Naïve Forecast with the Aid of the Diebold-Mariano Test

The French mathematician Pierre Simon Laplace (1814) introduced the principle of indifference (also known as the principle of insufficient reason) into the literature: A black box emits a figure x, and then the subject is requested to forecast which figure the black box will emit next. In view of the subject's complete lack of knowledge regarding the processes going on in the black box, it is not possible to give a single reason why the next figure should be larger than x. They can also not give a single reason why the next figure should be smaller than x. The only thing which an unknowing but sensible person can do is to forecast the figure x again for the future. In this way, a naïve forecast (everything remains the same) is understandable as long as one has no insight into the processes which lead to the figures which need to be forecast. Ever since it was identified, the naïve forecast has been considered to be the rock bottom in terms of forecast accuracy. Even a very rudimentary understanding of the processes at play should lead to better accuracy than that offered by a naïve forecast.

Simple measurements of forecast quality (such as mean absolute error or 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 avoided by using the Diebold-Mariano test (Diebold and Mariano, 1995). To do so, we calculate the mean squared error (MSE) 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:

$$DM = \frac{\frac{1}{T}\sum(V(P_{t1})-V(P_{t2}))}{\sqrt{\hat{\gamma}\,d/T}}$$

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 a clearly better result. The numerator is the mean deviation between the loss functions V of the two forecast 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 functions. In the case of large samples, this test value is asymptotically normally distributed.

3.2 Sign Accuracy Test

The sign accuracy test (Merton, 1981; Henriksson and Merton, 1981) is another widespread tool for evaluating forecasts. In this procedure, the extent of a forecasted change is not the issue. It only examines whether the general direction of the forecasts (rising or falling) is correct. The forecasts are then entered into a 2x2 matrix (Table 2).

Table 2: 2x2 Contingency Table

	Actual Event: Interest Rates Rise	Actual Event: Interest Rates Fall	Σ
Forecast: Interest Rates Rise	N_{11}	N_{12}	N_1
Forecast: Interest Rates Fall	N_{21}	N_{22}	N_2
Σ	N_1	N_2	N

On the one hand, a differentiation is made between whether an interest rate increase or an interest rate fall was forecast; on the other hand, a differentiation is also made between whether an interest rate rise or an interest rate fall has actually occurred. The principal diagonal in the 2x2 matrix (N_{11} and N_{22}) indicates the forecasts which are correct regarding the trend direction. The secondary diagonal (N_{12} and N_{21}) indicates the forecasts which are incorrect regarding the trend direction. A chi squared test is now applied to examine whether the distribution frequency of the four fields is significantly different from a random walk forecast (cf. Diebold and Lopez, 1996; Joutz and Stekler, 2000). If this is the case, a comparison between the number of observations in the principal diagonals and the secondary diagonals must be carried out to establish whether the forecasts are significantly better or significantly worse than a random walk forecast.

3.3 Unbiasedness Test

The unbiasedness test using the Mincer-Zarnowitz regression (Mincer and Zarnowitz, 1969) can check whether the forecast errors are systematic. According to the theory of rational expectations, this should not be the case. The Mincer-Zarnowitz regression takes the following form:

$$A_t = \alpha + \beta\, P_t + u_t$$

A_t = event which has actually occurred (dependent variable)

α = constant

P_t = forecast of the actual event at the moment in time t

β = coefficient of the respective forecasts

u_t = error term at the moment 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. In addition, the error term u_t may not be autocorrelated.

Forecasts are considered unbiased when, with a low probability of error, the joint hypothesis of α = 0 and β = 1 does not have to be rejected. This is checked by using the Wald test. A further condition is the absence of autocorrelations in the value of the error term u_t, which is examined with the Durbin-Watson test. If, according to these criteria, a forecast time series is based on rational expectations, Granger and Newbold (1973) argue that this by no means signifies that the forecasts are perfect. They merely do not exhibit systematic errors.

3.4 Topically-Orientated Trend Adjustment

In order to answer the question of whether forecasters have oriented themselves towards current levels when drawing up interest rate forecasts, the TOTA coefficient is used as a statistical benchmark (Andres and Spiwoks, 1999). Topically-orientated trend adjustment (TOTA) is present when forecasts reflect the present more strongly than the future. In the most unfavorable case, the future-oriented character of such forecasts may be lost entirely.

The TOTA coefficient is the quotient of two coefficients of determination (R^2_A and R^2_B). The R^2_A measures the correlation between the forecasts at the time of their validity and the actual events. The R^2_B measures the correlation between the forecasts at the time of their appearance and the actual events. The TOTA coefficient takes the following form:

$$TOTA\ coefficient = \frac{R^2_{\text{forecasts (validity date); actual events}}}{R^2_{\text{forecasts (issue date); actual events}}} = \frac{R^2_A}{R^2_B}$$

If the TOTA coefficient has a value of < 1, topically-orientated trend adjustment is given, and forecasts reflect the present more strongly than the future.

The TOTA coefficient and the unbiasedness test are closely related. If a forecast time series is characterized by the phenomenon of topically-orientated trend adjustment, the forecast error u_t is normally not randomly distributed (cf. Spiwoks, Bedke and Hein, 2010). Forecast time series which have a TOTA coefficient of < 1 are therefore normally biased.

4 Results

510 of the 532 forecasts analyzed have a TOTA coefficient of < 1 (see Tables 3-19). 95.9% of all the forecast time series analyzed are therefore characterized by the phenomenon of topically-orientated trend adjustments. If interest rates rise, expectations regarding future interest rates will therefore normally be revised upwards. If interest rates fall, expectations regarding future interest rates will therefore usually be revised downwards. In this way, the forecast

time series ultimately reflect current interest rates more strongly than future ones. Expressed more pointedly, it could be said that the experts are forecasting the present rather than the future. This is consistent with the results of earlier studies. In an analysis of 1,182 forecast time series of the G7 countries and five other European countries, a total of 98.5% of all forecast time series studied exhibited a topically-orientated trend adjustment (see Spiwoks, Gubaydullina and Hein, 2015).

These sobering findings are also reflected in the unbiasedness test. 529 of the 532 forecasts analyzed exhibit bias (see Tables 3-19). In 99.4% of all forecast time series studied, either α differs significantly from 0, or β differs significantly from 1, or the error term u_t proves to be autocorrelated.

Even unbiased forecasts can exhibit dramatic forecasting errors. The term unbiased merely states that forecasting errors are not of a systematic nature. A systematic forecasting error is, for example, a continuous over – or underestimation of the subject of the forecast ($\alpha \neq 0$). A different kind of systematic forecasting error is present when small actual events are constantly overestimated (or underestimated), and major actual events are constantly underestimated (or overestimated) ($\beta \neq 1$). Systematic forecasting errors are also present when the error term u_t reveals a pattern. This is usually the case when topically-orientated trend adjustment is present (cf. Spiwoks, Bedke and Hein 2010). However, other systematic forecasting errors can also lead to the error term u_t proving to be autocorrelated. Biased forecast time series are thus a reflection of systematic errors in drawing up the forecasts. This is true for 99.4% of all the forecasts we considered.

An expert's forecast can be viewed as largely worthless if it cannot bear comparison with the respective naïve forecast. A naïve forecast requires no specialist knowledge and is available free of charge to everyone at any time. One should, however, expect that forecasts made by highly-paid financial market experts are more exact than naïve forecasts. In many of the forecast subjects and forecast horizons examined here, the experts' forecasts – compared to the mean squared forecast error – are indeed more precise than naïve forecasts.[6] A total of 175 out of 532 forecast time series (32.9%) exhibit lower mean squared forecast errors than the respective naïve forecasts. However, the Diebold-Mariano test shows that only 19 out of 532 forecast time series (3.6%) contain significantly better forecasting results than naïve forecasts. The experts who forecast the prime lending rate in Hong Kong are particularly successful. 14 out of 30 forecast time series (46.7%) predict the interest rate trend significantly better than

---

[6] In the forecasts of the prime lending rate in Hong Kong, it can be seen that 26 out of 30 forecast time series (86.7%) were superior to the naïve forecast. In the case of the forecasts of the 3-month rate in Hong Kong, 24 out of 38 forecast time series (63.2%) were superior to a naïve forecast, while in the forecasts of the 3-month rate in India, at least the forecasts with a 13 month forecast horizon were highly successful: 9 out of 12 forecast time series (75%) were more precise than the corresponding naïve forecast. Among the forecasts of 10-year interest rates in Indonesia, the forecasts with a horizon of 13 months were once again very successful. 9 out of 12 forecast time series (75%) are more exact than a naïve forecast. Forecasts of the base lending rate in Malaysia were more successful than a naïve forecast in 18 out of 30 cases (60%), which is also the case for forecasts of 3-month interest rates in Malaysia in 19 out of 36 cases (52.8%). The forecasts of the 3-month rate in New Zealand were more precise than a naïve forecast in 25 out of 36 cases (69.4%). Among the forecasts of the 3-month rate in Thailand with a forecast horizon of 13 months, 9 out of 13 forecast time series (69.2%) were superior to the naïve forecast.

a naïve forecast (Table 6). Apart from this there are only five individual cases in which the time series of expert forecasts are significantly more precise than the time series of the respective naïve forecasts.

The sign accuracy test merely reveals whether forecasts were in the right direction (rising or falling). For the sign accuracy test, however, it is completely irrelevant whether forecasts predict the extent of future trends. The findings here are surprisingly favorable. In 248 out of 513 forecast time series (48.3%), the future trend (rising or falling interest rates) has been grasped significantly better than by a random walk forecast (see Tables 3-19). This is also a remarkable success in comparison to the findings of many previous studies. For example, Spiwoks, Bedke and Hein (2008) established a success rate of only 19.9% among US interest rate forecasts.

In the case of Australian 3-month interest rates with a forecast horizon of four months, 13 out of 21 forecast time series (61.9%) were significantly better in predicting the future trend direction (rising or falling) than a random walk forecast (Table 4). The forecasts for the base lending rate in China are very conspicuous: 29 out of 30 forecast time series (96.7%) predict the future interest rate trend significantly better than a random walk forecast (Table 5). This result is even surpassed by forecasts for the prime lending rate in Hong Kong. All 30 forecast time series (100%) reflect the future interest rate trend significantly more precisely than a random walk forecast (Table 6). The forecasts for three-month interest rates in Hong Kong, with a 13 month forecast horizon, are also very successful. 14 out of 19 forecast time series (73.7%) predict the interest rate trend significantly better than a random walk forecast (Table 7). Forecasts for the three-month rate in India are equally successful. In 17 out of 24 forecast time series (70.8%), the future trend (rising or falling interest rates) is reflected significantly better than by a random walk forecast (Table 9). The base lending rate in Malaysia is also forecasted successfully: 23 out of 28 forecast time series (82.1%) predict the future interest rate trend significantly better than a random walk forecast (Table 11). The forecasts for the three-month rate in New Zealand similarly predict the future interest rate trend significantly better (rising or falling) in 25 out of 36 cases (69.4%) than a random walk forecast (Table 14). Among the forecasts for three-month interest rates in Thailand, it is particularly those with a forecast horizon of four months that are successful. 10 out of 13 forecast time series (76.9%) predict the future trend significantly more precisely than a random walk forecast (Table 19).

In the case of 19 out of 532 forecast time series, the sign accuracy test could not be carried out, because frequencies of < 1 occur in one or several fields of the 2x2 contingency table. In these cases, however, the chi squared distribution is no longer a suitable test statistic (see, for example, Spiwoks, Bedke and Hein, 2009).

Overall, it can be stated that forecasting three-month interest rates is considerably easier than ten-year interest rates. Only 15.3% of the forecast time series on 10-year rates (Australia, India, Indonesia, New Zealand, Taiwan) predict the future trend (rising or falling interest rates) significantly more precisely than a random walk forecast, whereas in the case of three-month interest rates (Australia, Hong Kong, India, Malaysia, New Zealand, Singapore, Thailand) the figure is 57.1%. This coincides with the findings which have been obtained in other parts of the world. For example, in the case of US interest rate forecasts, Spiwoks, Bedke and Hein

(2008) showed that only 8.8% of all forecast time series on 10-year interest rates were significantly more successful than a random walk forecast, while in the case of three-month interest rates the figure was 30.9%.

The interest rates for short maturities are influenced considerably more by the actions of central banks than the interest rates for long maturities. In addition, central banks frequently provide an outlook on their future base rate policies. It can be that careful observation of central bank policy benefits forecasts of three-month interest rates, but not those for ten-year interest rates (cf. Spiwoks, Bedke and Hein 2008, p. 376). That would explain the variations in the success of forecasts.

Table 3: Results of the Measurement of Forecast Quality for Australia (10-Year Government Bond Yield)

| | | Forecast Horizon 4 Months | | | | | Unbiasedness | | | Forecast Horizon 13 Months | | | | | Unbiasedness | |
| | | | DM Test | | Sign Acc. Test | | F Test | DW | | DM Test | | Sign Acc. Test | | F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
ANZ	440	0.813	-	0.004	o	0.163	0.000	0.000	0.406	-	0.034	o	0.388	0.000	0.000
BIS Shrapnel	442	0.837	-	0.000	o	0.821	0.000	0.000	0.559	-	0.000	-	0.048	0.000	0.000
BT Financial Group	450	0.808	-	0.025	o	0.803	0.000	0.000	0.431	-	0.024	o	0.108	0.000	0.000
Centre of Policy St.	205	0.461	-	0.011	o	0.994	0.000	0.000	0.022	-	0.049	o	0.743	0.000	0.000
Citigroup	245	0.796	-	0.002	o	0.686	0.000	0.000	0.296	-	0.079	o	0.334	0.000	0.000
Commonwealth B.	444	0.785	-	0.000	-	0.029	0.000	0.000	0.401	-	0.015	o	0.642	0.000	0.000
Deloitte Acc. Econ.	428	0.806	-	0.017	o	0.290	0.000	0.000	0.424	o	0.111	-	0.011	0.000	0.000
Deutsche Bank	124	0.349	o	0.148	o	0.856	0.000	0.000	0.256	-	0.075	o	0.107	0.000	0.000
Goldman Sachs	228	0.743	-	0.009	o	0.542	0.000	0.000	0.415	o	0.183	o	0.564	0.000	0.000
HSBC	286	0.814	-	0.096	o	0.277	0.000	0.000	0.623	-	0.031	o	0.178	0.000	0.000
JPMorgan Chase	403	0.760	-	0.000	-	0.094	0.000	0.000	0.361	-	0.000	o	0.198	0.000	0.000
Macquarie	386	0.752	-	0.035	-	0.064	0.000	0.000	0.363	-	0.001	-	0.025	0.000	0.000
Merrill Lynch	300	0.847	-	0.018	o	0.889	0.000	0.000	0.570	o	0.242	o	0.266	0.000	0.000
Moody's Analytics	206	0.755	-	0.006	o	0.952	0.000	0.000	0.341	o	0.106	o	0.924	0.000	0.000
Nation. Australia B.	411	0.825	-	0.011	o	0.713	0.000	0.000	0.499	-	0.018	o	0.943	0.000	0.000
Nomura	328	0.587	-	0.023	o	0.675	0.000	0.000	0.137	-	0.024	-	0.011	0.000	0.000
Royal B. of Canada	272	0.771	o	0.249	o	0.114	0.000	0.000	0.298	o	0.107	o	0.576	0.000	0.000
Suncorp	212	0.436	-	0.000	o	0.489	0.000	0.000	0.038	o	0.155	+	0.057	0.000	0.000
UBS	449	0.791	-	0.005	o	0.973	0.000	0.000	0.382	-	0.023	-	0.006	0.000	0.000
Westpac	450	0.798	-	0.000	+	0.031	0.000	0.000	0.455	-	0.042	o	0.316	0.000	0.000
Consensus Forec.	504	0.806	-	0.004	o	0.887	0.000	0.000	0.430	-	0.031	o	0.323	0.000	0.000

\# = number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; Sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 4: Results of the Measurement of Forecast Quality for Australia (3-Month Interest Rates)

| | | | Forecast Horizon 4 Months | | | | Unbiasedness | | | | Forecast Horizon 13 Months | | | | Unbiasedness | |
| | | | DM Test | | Sign Acc. Test | | F Test | DW | | DM Test | | Sign Acc. Test | | F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
ANZ	468	0.904	o	0.471	+	0.000	0.000	0.000	0.377	o	0.180	+	0.035	0.000	0.000
BIS Shrapnel	465	0.897	o	0.159	+	0.001	0.000	0.000	0.490	o	0.215	+	0.001	0.000	0.000
BT Financial Group	468	0.886	o	0.432	+	0.000	0.109	0.000	0.427	o	0.847	+	0.002	0.000	0.000
Centre of Policy St.	219	0.780	o	0.393	o	0.107	0.001	0.000	0.015	-	0.096	o	0.663	0.000	0.000
Citigroup	276	0.854	o	0.889	o	0.156	0.002	0.000	0.239	o	0.235	o	0.379	0.000	0.000
Commonwealth B.	470	0.912	o	0.759	o	0.290	0.000	0.000	0.378	o	0.190	o	0.372	0.000	0.000
Deloitte Acc. Econ.	450	0.874	o	0.592	o	0.205	0.000	0.000	0.430	o	0.407	o	0.063	0.000	0.000
Deutsche Bank	124	0.560	o	0.867	o	0.358	0.000	0.000	0.044	o	0.089	o	0.839	0.000	0.000
Goldman Sachs	227	0.913	o	0.162	+	0.000	0.006	0.000	0.457	o	0.351	+	0.000	0.409	0.000
HSBC	272	0.953	o	0.502	o	0.109	0.000	0.000	0.713	o	0.172	+	0.028	0.000	0.000
JPMorgan Chase	410	0.877	o	0.450	+	0.008	0.001	0.000	0.387	o	0.902	o	0.387	0.000	0.000
Macquarie	411	0.900	o	0.537	+	0.005	0.000	0.000	0.368	o	0.388	o	0.482	0.000	0.000
Merrill Lynch	325	0.914	o	0.633	+	0.006	0.008	0.000	0.523	o	0.723	o	0.892	0.000	0.000
Moody's Analytics	188	0.906	o	0.457	+	0.002	0.028	0.000	0.364	o	0.779	o	0.660	0.000	0.000
Nation. Australia B.	432	0.896	o	0.366	+	0.046	0.843	0.000	0.452	o	0.970	o	0.066	0.001	0.000
Nomura	352	0.817	o	0.627	o	0.887	0.000	0.000	0.140	o	0.573	o	0.087	0.000	0.000
Royal B. of Canada	272	0.865	o	0.121	+	0.000	0.000	0.000	0.383	o	0.612	o	0.366	0.036	0.000
Suncorp	212	0.619	-	0.083	o	0.922	0.796	0.000	0.044	o	0.668	o	0.646	0.000	0.000
UBS	472	0.874	o	0.707	+	0.031	0.006	0.000	0.363	o	0.576	o	0.401	0.000	0.000
Westpac	470	0.930	o	0.506	+	0.008	0.000	0.000	0.503	o	0.234	+	0.002	0.000	0.000
Consensus Forec.	504	0.897	o	0.308	+	0.000	0.001	0.000	0.427	o	0.737	o	0.206	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 5: Results of the Measurement of Forecast Quality for China (1-Year Base Lending Rate)

| | | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Bank of China	258	0.671	o	0.317	+	0.000	0.000	0.000	0.037	o	0.623	+	0.000	0.000	0.000
Barclays Capital	101	0.841	o	0.733	o	0.317	0.000	0.000	0.030	o	0.386	+	0.000	0.000	0.000
BNP Paribas	108	0.888	o	0.830	+	0.009	0.000	0.000	0.441	o	0.343	+	0.000	0.000	0.000
Citigroup	100	0.599	o	0.229	+	0.001	0.000	0.047	0.007	o	0.179	+	0.000	0.000	0.000
Daiwa Capital	115	0.897	o	0.766	+	0.000	0.000	0.952	0.675	o	0.359	+	0.000	0.000	0.000
Deutsche Bank	158	0.702	o	0.240	+	0.036	0.002	0.000	0.000	o	0.445	+	0.000	0.000	0.000
Hang Seng Bank	107	0.918	o	0.532	+	0.001	0.001	0.000	0.547	o	0.885	+	0.590	0.000	0.000
HSBC Economics	212	0.665	o	0.332	+	0.000	0.000	0.000	0.045	o	0.530	+	0.000	0.000	0.000
IHS Economics	278	0.735	o	0.408	+	0.000	0.000	0.000	0.026	o	0.441	+	0.000	0.000	0.000
ING	216	0.574	o	0.298	+	0.002	0.000	0.000	0.036	o	0.889	+	0.000	0.000	0.000
JPMorgan Chase	96	0.580	o	0.271	+	0.000	0.001	0.001	0.000	o	0.324	+	0.002	0.000	0.098
Morgan Stanley	93	0.808	o	0.632	+	0.008	0.000	0.000	0.013	o	0.895	+	0.000	0.000	0.000
Nomura	158	1.017	o	0.860	+	0.000	0.001	0.000	1.432	o	0.751	+	0.000	0.000	0.000
Oxford Economics	232	0.695	-	0.089	+	0.000	0.000	0.000	0.153	o	0.618	+	0.000	0.000	0.000
Consensus Forec.	300	0.718	o	0.564	+	0.000	0.000	0.000	0.040	o	0.738	+	0.000	0.000	0.000

\# = number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 6: Results of the Measurement of Forecast Quality for Hong Kong (Prime Lending Rate)

| | | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Bank of China	334	0.967	+	0.025	+	0.000	0.026	0.000	0.574	+	0.058	+	0.000	0.000	0.000
Bank of East Asia	483	0.986	+	0.024	+	0.000	0.003	0.000	0.738	o	0.257	+	0.000	0.000	0.000
C. Pacific-Yamaichi	168	0.985	o	0.264	+	0.000	0.004	0.000	0.287	o	0.131	+	0.012	0.000	0.000
Credit Suisse	168	0.957	o	0.211	+	0.000	0.010	0.000	0.479	o	0.203	+	0.020	0.000	0.000
Daiwa Research I.	273	0.956	+	0.044	+	0.000	0.121	0.000	0.488	+	0.076	+	0.000	0.000	0.000
Deutsche Bank	264	0.967	+	0.031	+	0.000	0.159	0.000	0.633	o	0.418	+	0.000	0.000	0.000
FAZ Institute	132	0.939	o	0.208	+	0.000	0.829	0.000	0.449	o	0.161	+	0.001	0.000	0.000
Goldman Sachs	156	0.972	o	0.703	+	0.001	0.011	0.000	0.785	o	0.697	+	0.000	0.000	0.000
Hang Seng Bank	363	0.994	+	0.021	+	0.000	0.002	0.000	0.778	+	0.047	+	0.000	0.006	0.000
HSBC Economics	341	0.999	o	0.682	+	0.000	0.083	0.000	0.787	o	0.197	+	0.000	0.349	0.000
IHS Economics	288	0.957	+	0.075	+	0.000	0.003	0.000	0.527	o	0.201	+	0.000	0.000	0.000
Sakura Institute	139	0.378	o	0.946	+	0.000	0.169	0.000	0.153	+	0.009	+	0.000	0.000	0.000
S. Chartered Bank	209	0.970	+	0.036	+	0.000	0.023	0.000	0.441	o	0.668	+	0.000	0.000	0.000
UBS	132	0.930	+	0.098	+	0.000	0.863	0.001	0.253	o	0.408	+	0.045	0.455	0.000
Consensus Forec.	504	0.975	+	0.023	+	0.000	0.042	0.000	0.720	+	0.043	+	0.000	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 7: Results of the Measurement of Forecast Quality for Hong Kong (3-Month Interest Rates)

| | | | Forecast Horizon 4 Months | | | | Unbiasedness | | | | Forecast Horizon 13 Months | | | Unbiasedness | |
| | | | DM Test | | Sign Acc. Test | | F Test | DW | | DM Test | | Sign Acc. Test | | F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Bank of China	336	0.923	+	0.088	+	0.003	0.080	0.000	0.760	o	0.152	+	0.000	0.000	0.000
Bank of East Asia	484	0.906	o	0.121	+	0.000	0.010	0.000	0.804	o	0.163	+	0.001	0.000	0.000
Citigroup	339	0.967	o	0.125	+	0.000	0.848	0.000	0.768	o	0.390	+	0.001	0.000	0.000
C. Pacific-Yamaichi	168	0.981	o	0.628	+	0.000	0.000	0.000	0.374	o	0.317	+	0.001	0.000	0.000
Credit Suisse	185	0.887	o	0.231	o	0.091	0.292	0.000	0.573	o	0.577	o	0.284	0.000	0.000
Daiwa Research	275	0.818	+	0.096	o	0.191	0.064	0.000	0.481	o	0.418	o	0.301	0.000	0.000
Deutsche Bank	297	0.921	o	0.900	o	0.067	0.012	0.000	0.744	o	0.736	+	0.014	0.000	0.000
FAZ Institute	132	0.921	o	0.456	o	0.392	0.061	0.000	0.374	o	0.931	+	0.030	0.000	0.000
Goldman Sachs	375	0.862	o	0.214	+	0.000	0.081	0.000	0.795	o	0.223	+	0.000	0.001	0.000
Hang Seng Bank	363	0.913	o	0.143	+	0.023	0.595	0.000	0.832	+	0.041	+	0.000	0.011	0.000
HSBC	342	0.962	o	0.318	o	0.244	0.082	0.000	0.827	o	0.182	+	0.000	0.002	0.000
ING	405	0.920	o	0.202	+	0.000	0.339	0.000	0.794	o	0.368	+	0.000	0.000	0.000
Morgan Stanley	100	1.013	o	0.411	o	0.401	0.240	0.761	0.818	o	0.744	+	0.008	0.000	0.000
Nomura	196	0.947	o	0.251	o	0.474	0.000	0.000	0.868	o	0.336	+	0.001	0.000	0.000
Sakura Institute	139	0.126	o	0.291	o	0.098	0.123	0.001	0.002	o	0.301	o	0.077	0.044	0.000
Societe Generale	117	0.834	o	0.464	o	0.184	0.024	0.094	0.676	o	0.327	o	0.149	0.004	0.000
S. Chartered Bank	212	0.782	o	0.415	o	0.521	0.000	0.002	0.392	o	0.897	+	0.013	0.000	0.000
UBS	131	0.808	o	0.638	o	0.674	0.001	0.001	0.026	o	0.767	o	0.510	0.000	0.000
Consensus Forec.	504	0.909	o	0.154	+	0.000	0.001	0.000	0.778	o	0.140	+	0.000	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 8: Results of the Measurement of Forecast Quality for India (10-Year Government Bond Yield)

| | | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	190	0.182	o	0.110	O	0.273	0.007	0.000	0.025	o	0.725	+	0.001	0.004	0.000
Confed of Indian I.	144	0.949	-	0.037	o	0.073	0.000	0.032	0.684	o	0.563	o	0.093	0.000	0.000
Deutsche Bank	101	0.932	-	0.011	+	0.020	0.090	0.000	0.907	o	0.329	+	0.030	0.008	0.000
Dresdner Bank	84	0.747	o	0.919	+	0.011	0.021	0.790	0.653	o	0.287	+	0.037	0.183	0.829
FERI	156	0.389	-	0.058	o	0.270	0.000	0.000	0.003	o	0.245	o	0.123	0.000	0.000
Hindustan Lever	176	0.897	-	0.063	o	0.601	0.000	0.000	0.718	o	0.112	o	0.500	0.000	0.000
HSBC Securities	272	0.849	o	0.126	o	0.080	0.015	0.000	0.812	o	0.299	+	0.000	0.000	0.000
HIS Economics	186	0.607	o	0.170	+	0.028	0.076	0.000	0.317	o	0.140	-	0.046	0.000	0.000
NCAER	214	0.881	-	0.005	o	0.214	0.000	0.000	0.712	-	0.003	o	0.688	0.000	0.000
Nomura	224	0.951	o	0.171	o	0.106	0.000	0.000	0.951	o	0.537	+	0.000	0.000	0.000
Tata Services	327	0.922	-	0.073	+	0.001	0.033	0.000	0.813	o	0.163	+	0.000	0.000	0.000
UBS	138	0.905	o	0.178	o	0.128	0.000	0.000	0.730	o	0.518	+	0.013	0.000	0.000
Consensus Forec.	504	0.934	o	0.167	+	0.001	0.000	0.000	0.843	o	0.944	+	0.000	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 9: Results of the Measurement of Forecast Quality for India (3-Month Interest Rates)

| | | | Forecast Horizon 4 Months | | | | Unbiasedness | | | | Forecast Horizon 13 Months | | | | Unbiasedness | |
| | | | DM Test | | Sign Acc. Test | | F Test | DW | | DM Test | | Sign Acc. Test | | F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	221	0.862	o	0.316	o	0.558	0.000	0.000	0.104	o	0.669	+	0.001	0.000	0.000
Confed of Indian I.	166	0.908	o	0.621	+	0.001	0.000	0.009	0.656	o	0.624	+	0.028	0.000	0.002
Deutsche Bank	151	0.905	o	0.304	+	0.011	0.109	0.000	0.443	o	0.413	+	0.000	0.000	0.000
Dresdner Bank	232	0.808	NA	NA	+	0.010	0.001	0.000	0.373	o	0.296	+	0.000	0.018	0.000
Goldman Sachs	98	0.113	o	0.389	+	0.026	0.002	0.000	0.000	o	0.553	o	0.835	0.000	0.000
Hindustan Lever	168	0.864	NA	NA	+	0.011	0.000	0.001	0.752	-	0.001	+	0.000	0.000	0.002
HSBC Securities	234	0.502	o	0.172	o	0.585	0.000	0.000	0.167	o	0.513	+	0.000	0.000	0.000
JPMorgan	108	0.484	o	0.315	o	0.793	0.000	0.707	0.452	o	0.249	+	0.001	0.004	0.594
NCAER	212	0.809	o	0.335	+	0.000	0.006	0.000	0.336	o	0.927	o	0.090	0.000	0.000
Tata Services	325	0.854	o	0.487	+	0.000	0.000	0.000	0.180	o	0.870	+	0.000	0.000	0.000
UBS	136	0.666	o	0.158	o	0.200	0.001	0.000	0.049	o	0.186	o	0.133	0.122	0.002
Consensus Forec.	504	0.799	o	0.393	+	0.002	0.001	0.000	0.223	o	0.502	+	0.000	0.000	0.000

\# = number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 10: Results of the Measurement of Forecast Quality for India (10-Year Government Bond Yield)

| | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Bahana Securities	194	0.916	o	0.990	o	0.108	0.000	0.000	0.569	o	0.553	-	0.033	0.000	0.000
Bank Danamon	98	0.809	NA	NA	o	0.301	0.000	0.023	0.464	o	0.367	+	0.043	0.000	0.000
Castle Asia	108	0.787	-	0.034	o	0.433	0.360	0.000	0.440	o	0.513	+	0.003	0.000	0.000
Citigroup	228	0.912	o	0.505	o	0.273	0.000	0.615	0.443	o	0.707	o	0.722	0.000	0.000
Danareksa S.	375	1.001	NA	NA	o	0.422	0.000	0.000	0.844	o	0.273	o	0.317	0.156	0.000
GK Goh	110	1.107	o	0.347	o	0.053	0.276	0.000	0.969	o	0.864	o	0.542	0.000	0.000
HSBC Economics	272	0.916	o	0.270	o	0.983	0.000	0.000	0.590	o	0.281	o	0.601	0.000	0.000
ING	199	0.937	o	0.353	o	0.191	0.008	0.000	0.747	o	0.519	o	0.983	0.000	0.000
Nomura	134	0.867	o	0.338	o	0.866	0.000	0.095	0.419	o	0.308	-	0.004	0.000	0.000
Societe Generale	124	0.822	-	0.002	NA	NA	0.450	0.000	0.653	o	0.222	NA	NA	0.031	0.000
S. Chartered Bank	174	0.861	o	0.111	o	0.676	0.144	0.000	0.358	o	0.660	o	0.971	0.018	0.000
Consensus Forec.	504	0.926	o	0.265	o	0.884	0.000	0.000	0.590	o	0.501	o	0.547	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 11: Results of the Measurement of Forecast Quality for Malaysia (Base Lending Rate)

| | | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
AMSecurities	212	0.890	o	0.226	+	0.000	0.000	0.000	0.493	o	0.318	o	0.309	0.000	0.000
CIBD-CIMB	157	1.002	+	0.096	+	0.003	0.001	0.000	0.612	o	0.272	+	0.017	0.000	0.000
Citigroup	94	0.848	o	0.305	+	0.007	0.000	0.000	0.105	o	0.821	+	0.013	0.000	0.000
Deutsche Bank	92	0.996	o	0.261	NA	NA	0.022	0.000	1.016	o	0.222	+	0.000	0.000	0.010
Goldman Sachs	118	0.792	-	0.049	+	0.000	0.000	0.000	0.388	o	0.301	+	0.000	0.000	0.000
HSBC Securities	149	0.551	-	0.056	o	0.178	0.000	0.000	0.005	o	0.441	+	0.001	0.000	0.000
JM Sassoon	150	0.918	o	0.218	+	0.000	0.055	0.000	0.746	o	0.158	+	0.001	0.270	0.000
Kanega Research	109	0.954	o	0.131	+	0.009	0.014	0.000	0.142	o	0.155	+	0.012	0.000	0.000
Kay Hian Research	261	0.951	o	0.249	+	0.001	0.495	0.000	0.918	o	0.231	+	0.001	0.368	0.000
Maybank	145	0.791	o	0.390	NA	NA	0.082	0.000	0.037	-	0.086	o	0.201	0.000	0.000
MIER	276	0.862	o	0.277	+	0.000	0.095	0.000	0.636	o	0.475	o	0.697	0.204	0.000
RHB Research	400	1.021	o	0.287	+	0.000	0.000	0.000	0.676	o	0.169	+	0.000	0.045	0.000
Societe Generale	104	0.911	o	0.216	+	0.000	0.000	0.000	1.102	o	0.800	+	0.008	0.000	0.000
S. Chartered Bank	165	1.051	o	0.240	+	0.000	0.003	0.000	0.654	o	0.555	o	0.059	0.000	0.000
Consensus Forec.	480	0.939	o	0.130	+	0.000	0.005	0.000	0.606	o	0.160	+	0.000	0.180	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 12: Results of the Measurement of Forecast Quality for Malaysia (3-Month Interest Rates)

| | | Forecast Horizon 4 Months | | | | | Unbiasedness | | | Forecast Horizon 13 Months | | | | Unbiasedness | |
| | | | DM Test | | Sign Acc. Test | | F Test | DW | | DM Test | | Sign Acc. Test | | F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
AMSecurities	211	0.896	o	0.261	o	0.051	0.706	0.000	0.364	o	0.400	o	0.161	0.000	0.000
Baring- ING	427	0.882	o	0.499	+	0.000	0.072	0.000	0.603	o	0.920	+	0.000	0.000	0.000
BofA-Merrill Lynch	111	1.050	o	0.643	o	0.068	0.052	0.000	0.924	o	0.650	o	0.536	0.000	0.000
CIBD-CIMB	156	0.962	o	0.622	o	0.350	0.028	0.000	0.588	o	0.184	o	0.148	0.000	0.000
Citigroup	383	0.923	o	0.100	+	0.004	0.011	0.000	0.534	o	0.474	+	0.002	0.000	0.000
Deutsche Bank	127	0.996	o	0.546	o	0.071	0.006	0.000	0.884	o	0.153	+	0.000	0.000	0.014
Goldman Sachs	464	0.877	-	0.081	o	0.663	0.000	0.000	0.573	NA	NA	+	0.017	0.000	0.000
HSBC Economics	195	0.819	-	0.051	o	0.060	0.000	0.000	0.138	o	0.479	+	0.004	0.000	0.000
JM Sassoon	150	0.917	o	0.180	+	0.000	0.472	0.000	0.689	o	0.248	+	0.002	0.064	0.000
Kanega Research	118	0.819	o	0.113	+	0.005	0.098	0.000	0.077	o	0.179	+	0.000	0.047	0.000
Kay Hian Research	120	0.841	o	0.575	+	0.001	0.009	0.001	0.497	o	0.276	o	0.940	0.000	0.000
Maybank	141	0.876	o	0.110	o	0.699	0.008	0.000	0.200	o	0.144	o	0.248	0.000	0.000
MIER	282	0.915	o	0.243	+	0.002	0.006	0.000	0.604	o	0.424	o	0.482	0.000	0.000
RHB Research	404	0.935	o	0.245	+	0.000	0.527	0.000	0.501	o	0.263	+	0.000	0.000	0.000
Societe Generale	118	1.019	o	0.938	+	0.000	0.436	0.001	1.540	o	0.810	o	0.095	0.000	0.000
S. Chartered Bank	238	0.963	o	0.237	+	0.001	0.028	0.000	0.581	o	0.445	o	0.238	0.025	0.000
UOB Kay Hian	220	0.908	o	0.274	o	0.189	0.654	0.000	0.352	o	0.294	+	0.007	0.000	0.000
Consensus Forec.	504	0.927	o	0.199	+	0.000	0.154	0.000	0.588	o	0.291	+	0.000	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 13: Results of the Measurement of Forecast Quality for New Zealand (10-Year Government Bond Yield)

			Forecast Horizon 4 Months							Forecast Horizon 13 Months					
			DM Test		Sign Acc. Test		Unbiasedness F Test	DW		DM Test		Sign Acc. Test		Unbiasedness F Test	DW
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
ANZ Bank	488	0.832	-	0.010	o	0.286	0.000	0.000	0.575	-	0.053	+	0.007	0.000	0.000
ASB Bank	274	0.798	-	0.011	o	0.849	0.000	0.000	0.393	-	0.072	o	0.630	0.000	0.000
Bank of NZ	480	0.779	-	0.038	o	0.806	0.000	0.000	0.426	-	0.047	o	0.423	0.000	0.000
BERL	340	0.483	-	0.024	o	0.951	0.000	0.000	0.231	-	0.086	o	0.355	0.043	0.000
Credit Suisse FB	114	0.297	o	0.263	o	0.602	0.002	0.000	0.239	o	0.983	o	0.638	0.213	0.000
Deutsche Bank NZ	468	0.831	-	0.006	+	0.032	0.000	0.000	0.481	-	0.001	-	0.004	0.000	0.000
First NZ Capital	348	0.830	-	0.000	o	0.988	0.000	0.000	0.565	o	0.145	o	0.198	0.000	0.000
Goldman Sachs NZ	144	0.516	-	0.059	o	0.955	0.043	0.000	0.093	o	0.308	-	0.015	0.000	0.000
HSBC Economics	216	0.867	-	0.003	o	0.550	0.000	0.000	0.490	-	0.053	o	0.129	0.000	0.000
Infometrics	498	0.762	-	0.002	o	0.469	0.000	0.000	0.426	-	0.021	o	0.549	0.000	0.000
JPMorgan Chase	331	0.754	-	0.000	o	0.444	0.019	0.000	0.361	-	0.035	o	0.833	0.000	0.000
Macquarie	172	0.776	-	0.003	o	0.271	0.000	0.000	0.341	-	0.050	o	0.472	0.000	0.000
National Bank NZ	212	0.465	o	0.663	o	0.107	0.529	0.000	0.000	o	0.180	o	0.455	0.000	0.000
NZIER	440	0.764	-	0.000	o	0.257	0.000	0.000	0.572	-	0.060	+	0.002	0.000	0.000
Ord Minnett S.	128	0.234	-	0.005	o	0.567	0.008	0.000	0.367	o	0.783	+	0.000	0.000	0.000
UBS	476	0.779	-	0.001	-	0.038	0.000	0.000	0.393	-	0.084	-	0.040	0.000	0.000
Westpac	482	0.799	-	0.003	o	0.663	0.000	0.000	0.497	-	0.092	o	0.359	0.000	0.000
Consensus Forec.	504	0.793	-	0.005	o	0.377	0.000	0.000	0.504	o	0.118	o	0.075	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 14: Results of the Measurement of Forecast Quality for New Zealand (3-Month Interest Rates)

		Forecast Horizon 4 Months						Forecast Horizon 13 Months						
			DM Test		Sign Acc. Test		Unbiasedness F Test	DW		DM Test		Sign Acc. Test	Unbiasedness F Test	DW
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
ANZ Bank	488	0.939	o	0.287	+	0.000	0.063	0.000	0.694	o	0.383	o	0.446	0.000	0.000
ASB Bank	274	0.998	o	0.243	+	0.000	0.018	0.000	0.684	o	0.654	o	0.434	0.000	0.000
Bank of NZ	480	0.937	o	0.152	+	0.000	0.068	0.000	0.758	o	0.341	+	0.003	0.000	0.000
BERL	340	0.770	o	0.275	o	0.173	0.000	0.000	0.403	o	0.213	o	0.128	0.038	0.000
Credit Suisse FB	114	0.742	o	0.215	+	0.012	0.490	0.000	0.042	o	0.236	+	0.000	0.043	0.000
Deutsche Bank NZ	468	0.947	+	0.076	+	0.000	0.064	0.000	0.630	o	0.497	+	0.007	0.000	0.000
First NZ Capital	348	0.960	o	0.161	+	0.000	0.124	0.000	0.637	o	0.394	o	0.299	0.000	0.000
Goldman Sachs NZ	146	0.825	o	0.265	+	0.000	0.000	0.000	0.079	o	0.464	+	0.008	0.000	0.000
HSBC Economics	198	0.972	o	0.294	+	0.001	0.000	0.000	0.805	o	0.419	+	0.038	0.000	0.000
Infometrics	498	0.925	o	0.411	+	0.000	0.000	0.000	0.472	o	0.969	o	0.363	0.000	0.000
JPMorgan Chase	327	0.959	o	0.328	o	0.071	0.115	0.000	0.435	o	0.861	o	0.743	0.000	0.000
Macquarie	176	1.002	o	0.320	+	0.001	0.021	0.000	0.806	o	0.324	o	0.379	0.000	0.000
National Bank NZ	212	0.807	o	0.130	+	0.000	0.124	0.000	0.270	o	0.361	+	0.000	0.000	0.000
NZIER	440	0.891	o	0.997	o	0.067	0.015	0.000	0.493	o	0.860	o	0.258	0.000	0.000
Ord Minnett S.	128	0.772	o	0.537	+	0.009	0.078	0.000	0.685	o	0.193	+	0.001	0.860	0.000
UBS	478	0.927	o	0.412	+	0.001	0.496	0.000	0.558	o	0.571	+	0.028	0.000	0.000
Westpac	482	0.946	o	0.137	+	0.000	0.041	0.000	0.542	o	0.611	+	0.039	0.006	0.000
Consensus Forec.	504	0.929	o	0.150	+	0.000	0.562	0.000	0.603	o	0.291	+	0.009	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 15: Results of the Measurement of Forecast Quality for Singapore (Prime Lending Rate)

| | | | Forecast Horizon 4 Months | | | | | | | | Forecast Horizon 13 Months | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	94	0.580	o	0.604	NA	NA	0.004	0.000	0.308	o	0.163	+	0.000	0.000	0.000
Credit Suisse	95	0.847	-	0.000	NA	NA	0.000	0.000	0.355	-	0.000	NA	NA	0.000	0.002
Deutsche Bank	164	0.996	-	0.000	NA	NA	0.000	0.000	0.920	-	0.000	NA	NA	0.000	0.000
Goldman Sachs	151	0.792	-	0.003	o	0.696	0.000	0.000	0.871	o	0.196	+	0.003	0.000	0.000
HSBC	322	1.090	-	0.000	NA	NA	0.000	0.000	2.542	-	0.000	NA	NA	0.000	0.000
JM Sassoon	164	1.144	o	0.457	o	0.119	0.000	0.000	2.295	o	0.956	-	0.030	0.000	0.000
Kay Hian Research	122	0.784	o	0.280	o	0.055	0.000	0.000	1.378	o	0.951	+	0.001	0.000	0.000
Morgan Stanley	95	0.818	-	0.000	NA	NA	0.000	0.000	0.704	-	0.010	o	0.338	0.000	0.000
Nomura	183	0.973	-	0.000	o	0.505	0.000	0.000	1.054	-	0.008	+	0.003	0.000	0.000
OCBC Bank	194	0.970	-	0.000	NA	NA	0.000	0.000	0.868	-	0.000	NA	NA	0.000	0.000
Sakura Institute	144	0.749	o	0.430	+	0.007	0.000	0.000	0.398	o	0.903	+	0.003	0.000	0.000
S. Chartered Bank	168	0.978	-	0.061	o	0.107	0.000	0.000	0.863	o	0.395	o	0.064	0.000	0.000
U. Overseas Bank	190	0.937	-	0.013	o	0.058	0.000	0.000	1.050	o	0.542	+	0.004	0.000	0.000
UOB Kay Hian	143	9.037	-	0.000	NA	NA	0.000	0.000	2.053	-	0.000	NA	NA	0.000	0.000
Consensus Forec.	392	0.946	-	0.000	o	0.090	0.000	0.000	0.749	-	0.002	-	0.039	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 16: Results of the Measurement of Forecast Quality for Singapore (3-Month Interest Rates)

		Forecast Horizon 4 Months							Forecast Horizon 13 Months						
			DM Test		Sign Acc. Test		Unbiasedness F Test	DW		DM Test		Sign Acc. Test		Unbiasedness F Test	DW
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	382	0.847	o	0.196	+	0.017	0.002	0.000	0.431	o	0.189	o	0.229	0.353	0.000
Credit Suisse	154	0.922	o	0.246	o	0.058	0.270	0.000	0.424	o	0.531	o	0.264	0.000	0.000
DBS Bank	208	0.892	o	0.554	+	0.001	0.000	0.000	0.379	o	0.349	-	0.032	0.000	0.000
Deutsche Bank	240	0.916	o	0.999	+	0.003	0.017	0.047	0.761	o	0.844	+	0.049	0.000	0.000
Goldman Sachs	454	0.818	o	0.111	+	0.007	0.000	0.000	0.522	o	0.557	+	0.007	0.000	0.000
HSBC	335	0.926	o	0.553	o	0.103	0.000	0.000	0.501	o	0.948	-	0.002	0.000	0.000
IHS	212	0.891	o	0.119	o	0.990	0.451	0.000	0.311	-	0.098	o	0.622	0.000	0.000
ING	279	0.697	o	0.319	+	0.028	0.024	0.000	0.097	o	0.442	+	0.011	0.000	0.000
JM Sassoon	166	0.541	o	0.494	+	0.041	0.149	0.163	0.064	o	0.495	o	0.914	0.626	0.000
Kay Hian Research	236	0.572	-	0.057	o	0.375	0.059	0.000	0.188	o	0.426	+	0.023	0.000	0.000
Merrill Lynch	127	0.767	o	0.313	o	0.855	0.470	0.027	0.371	o	0.229	o	0.874	0.000	0.001
Nomura	252	0.663	o	0.571	o	0.385	0.202	0.000	0.426	o	0.319	+	0.031	0.000	0.000
OCBC Bank	323	0.831	o	0.554	o	0.291	0.000	0.000	0.527	o	0.750	+	0.002	0.000	0.000
Sakura Institute	128	0.315	-	0.052	o	0.796	0.001	0.444	0.001	o	0.391	o	0.492	0.000	0.028
S. Chartered Bank	243	0.658	o	0.484	+	0.016	0.049	0.002	0.098	o	0.823	o	0.066	0.000	0.000
UBS	122	0.625	o	0.951	o	0.435	0.146	0.002	0.355	o	0.492	o	0.411	0.000	0.000
U. Overseas Bank	206	0.751	o	0.412	+	0.026	0.176	0.000	0.615	o	0.240	+	0.000	0.031	0.000
UOB Kay Hian	143	0.837	o	0.434	o	0.184	0.051	0.000	0.342	o	0.719	+	0.000	0.000	0.000
Consensus Forec.	504	0.770	o	0.511	+	0.000	0.000	0.000	0.436	o	0.403	+	0.002	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

Table 17: Results of the Measurement of Forecast Quality for South Korea (3-Year Government Bond Yield)

| | | | Forecast Horizon 4 Months | | | | | | | Forecast Horizon 13 Months | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Credit Suisse	112	0.295	-	0.032	o	0.790	0.009	0.000	0.403	-	0.021	o	0.340	0.000	0.000
Daewoo Securities	208	0.849	-	0.090	o	0.874	0.000	0.000	0.586	o	0.465	o	0.172	0.000	0.000
Daishin Economics	133	0.634	-	0.098	o	0.466	0.000	0.000	0.164	-	0.000	-	0.001	0.001	0.000
Dresdner Bank	179	0.724	-	0.002	o	0.277	0.000	0.000	0.643	-	0.006	o	0.473	0.000	0.000
HSBC Economics	94	0.838	-	0.009	o	0.848	0.002	0.000	0.634	o	0.463	o	0.108	0.000	0.000
Hyundai Securities	228	0.714	-	0.005	o	0.364	0.000	0.000	0.216	-	0.056	o	0.166	0.000	0.000
ING Baring	94	0.322	o	0.149	-	0.046	0.000	0.000	0.382	-	0.000	-	0.040	0.001	0.000
LG Group	211	0.856	-	0.056	o	0.283	0.000	0.000	0.338	-	0.001	-	0.001	0.000	0.000
Samsung ER	196	0.923	-	0.000	o	0.198	0.000	0.000	0.830	-	0.021	-	0.025	0.000	0.000
Sakura	143	0.581	-	0.013	o	0.252	0.000	0.000	0.020	-	0.031	o	0.107	0.372	0.000
Shinhan Securities	144	0.640	-	0.002	NA	NA	0.000	0.000	0.138	-	0.000	NA	NA	0.000	0.000
Societe Generale	92	0.841	-	0.000	o	0.204	0.000	0.000	0.573	o	0.725	o	0.123	0.000	0.002
UBS	101	0.912	-	0.018	o	0.073	0.001	0.000	0.820	-	0.038	o	0.781	0.000	0.005
Consensus Forec.	278	0.834	-	0.047	o	0.795	0.000	0.000	0.485	-	0.008	-	0.014	0.000	0.000

\# = number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 18: Results of the Measurement of Forecast Quality for Taiwan (10-Year Government Bond Yield)

| | | Forecast Horizon 4 Months | | | | | | | | Forecast Horizon 13 Months | | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	176	0.325	o	0.116	o	0.791	0.000	0.043	0.000	-	0.036	o	0.645	0.000	0.294
HSBC	192	0.629	-	0.004	o	0.681	0.000	0.000	0.127	-	0.094	o	0.163	0.000	0.000
IHS	234	0.645	-	0.023	-	0.034	0.000	0.000	0.222	-	0.018	o	0.222	0.000	0.000
ING	224	0.675	o	0.119	o	0.752	0.000	0.005	0.208	o	0.120	o	0.941	0.000	0.000
Nomura	167	0.635	-	0.004	o	0.444	0.001	0.000	0.001	-	0.032	o	0.391	0.000	0.000
Polaris Research	134	0.721	-	0.000	o	0.601	0.000	0.000	0.339	-	0.003	o	0.314	0.000	0.000
Taiwan Institute R.	126	0.719	o	0.227	+	0.024	0.000	0.000	0.222	o	0.102	NA	NA	0.000	0.000
Consensus Forec.	236	0.672	-	0.045	o	0.453	0.000	0.000	0.201	-	0.065	o	0.234	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test; NA = not available.

Table 19: Results of the Measurement of Forecast Quality for Thailand (3-Month Interest Rates)

| | | | Forecast Horizon 4 Months | | | | | | Forecast Horizon 13 Months | | | | | | |
| | | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW | | DM Test | | Sign Acc. Test | | Unbiasedness F Test | DW |
Institution	#	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value	TOTA	Res	P-Value	Res	P-Value	P-Value	P-Value
Citigroup	298	0.823	o	0.366	+	0.003	0.000	0.000	0.468	o	0.570	+	0.000	0.000	0.000
Deutsche Bank	82	0.972	o	0.762	+	0.006	0.064	0.000	0.797	o	0.320	o	0.908	0.298	0.000
Goldman Sachs	377	0.762	o	0.231	o	0.297	0.063	0.000	0.522	o	0.368	o	0.366	0.000	0.000
HSBC Economics	346	0.865	o	0.349	+	0.000	0.164	0.000	0.361	o	0.332	o	0.060	0.000	0.000
ING	400	0.820	o	0.328	+	0.006	0.481	0.000	0.344	o	0.297	+	0.000	0.000	0.000
Kasikornbank	390	0.761	o	0.148	+	0.001	0.000	0.000	0.391	o	0.401	+	0.007	0.000	0.000
Merrill Lynch	155	0.865	o	0.270	o	0.659	0.000	0.002	0.538	o	0.524	-	0.025	0.000	0.000
Morgan Stanley	85	0.963	o	0.379	o	0.804	0.194	0.027	0.131	o	0.509	o	0.095	0.030	0.095
Nomura	146	0.791	o	0.487	+	0.000	0.000	0.421	0.579	o	0.912	+	0.006	0.000	0.007
Phatra Thanakit S.	334	0.850	o	0.370	+	0.007	0.000	0.000	0.554	o	0.382	+	0.001	0.000	0.000
Siam C. Bank	175	0.899	o	0.254	+	0.006	0.000	0.008	0.725	o	0.391	o	0.506	0.000	0.000
S. Chartered Bank	206	0.841	o	0.838	+	0.000	0.000	0.000	0.398	o	0.474	+	0.047	0.000	0.000
Consensus Forec.	504	0.841	o	0.437	+	0.000	0.008	0.000	0.477	o	0.379	o	0.186	0.000	0.000

= number of observations; TOTA = TOTA coefficient; DM test = Diebold-Mariano test; Res = result; o = no significant result; - = significantly worse than a naïve or random walk forecast; + = significantly better than a naïve or random walk forecast; sign acc. test = sign accuracy test; unbiasedness = test for unbiasedness; DW = Durbin-Watson test.

A further factor is that it is obviously more difficult to forecast market interest rates than those which are set or controlled by governments. The base lending rate in China, the prime lending rate in Hong Kong, the base lending rate and the three-month interest rate in Malaysia as well as the prime lending rate in Singapore are set directly by the respective administration or – at least partly – managed by it. The success rate for the relevant forecast time series is relatively high: 75% of the forecast time series for these interest rates predict the future interest rate trend (rising or falling) significantly better than a random walk forecast.

5 Conclusion

We analyzed interest rate forecasts for the Asia-Pacific region in the period from 1990-2016. To do so, we examined individual interest rate forecasts from Australia, China, Hong Kong, India, Indonesia, Malaysia, New Zealand, Singapore, South Korea, Taiwan and Thailand. As a basis we used forecasting data which had been published in the journal Asia Pacific Consensus Forecasts on a monthly basis. We did not limit ourselves to the analysis of consensus forecasts, however: We also evaluated all of the forecast time series issues by banks, investment companies, consulting firms, associations and industrial companies. Overall we assessed 532 forecast time series with a total of 85,264 individual interest rate forecasts. The variety of procedures which we used to measure the quality of forecasts enabled us to create a comprehensive evaluation of forecasting performance in the Asia-Pacific region. We carried out a comparison to naïve forecasts. We examined the forecast time series for evidence of topically-orientated trend adjustments. In addition, we deployed the sign accuracy test and the unbiasedness test.

The results are very sobering in part. 95.9% of all forecast time series are characterized by the phenomenon of topically-oriented trend adjustments. This means that the overwhelming majority of all forecast time series reflect the present rather than the future. In total, 99.4% of all forecast time series proved to be biased. Given that topically-orientated trend adjustments usually lead to the error term u_t not being distributed randomly, the result of the unbiasedness test is not surprising.

Only a small proportion of the forecast time series (3.6%) reflected the future interest rate trend significantly more precisely than a naïve forecast. The only forecast whose success went beyond rare individual cases was that for the prime lending rate in Hong Kong. 46.7% of these forecast time series predict the future interest rate trend significantly better than a random walk forecast.

However, some of the results of the study are also surprisingly positive. The sign accuracy test reveals that in 248 out of 513 forecast time series (48.3%), the future trend (rising or falling interest rates) has been grasped significantly better than by a random walk forecast. In this context, at least part of the forecasts for Australia, China, Hong Kong, India, Malaysia, New Zealand, Singapore and Thailand proved to be particularly successful.

Overall it can be stated that – at least in some countries and for some forecast horizons – forecasts of future interest rate trends in the Asia-Pacific region are significantly more successful than those made in other parts of the world. This has consequences for portfolio management: Active portfolio management strategies have no prospects of success in many financial markets because the necessary forecasting competence is simply not there. However, this is different, for example, in the case of the Indian bond market. 61.5% of the forecast time series on the interest rates of Indian state bonds with ten years remaining to maturity (forecast horizon: 13 months) predict the future interest rate trend (rising or falling) significantly better than a random walk forecast. This should suffice in order to achieve systematic excess returns with active portfolio management strategies.

Acknowledgments

References

Albrecht, T. (2000): Zur Eignung professioneller Zinsprognosen als Entscheidungsgrundlage, Ein Vergleich der Zinsprognosen deutscher Banken mit der Zinserwartung „des Marktes", Sofia – Diskussionsbeiträge zur Institutionenanalyse, 00-7, 1-29.

Andres, P. / Spiwoks, M. (1999): Prognosequalitätsmatrix – Ein methodologischer Beitrag zur Beurteilung der Güte von Kapitalmarktprognosen, Jahrbücher für Nationalökonomie und Statistik, 219(5/6), 513-542.

Baghestani, H. (2005): Improving the Accuracy of Recent Survey Forecasts of the T-Bill Rate, Business Economics, 40, 36-40.

Baghestani, H. / Arzaghi, M. / Kaya, I. (2015): On the Accuracy of Blue Chip Forecasts of Interest Rates and Country Risk Premiums, Applied Economics, 47(2), 113-122.

Baghestani, H. / Danila, L. (2014): Interest Rate and Exchange Rate Forecasting in the Czech Republic: Do Analysts Know Better than a Random Walk?, Czech Journal of Economics and Finance, 64(4), 282-295.

Baghestani, H. / Jung, W. / Zuchegno, D. (2000): On the Information Content of Futures Market and Professional Forecasts of Interest Rates, Applied Financial Economics, 10, 679-684.

Baghestani, H. / Marchon, C. (2012): An Evaluation of Private Forecasts of Interest Rate Targets in Brazil, Economics Letters, 115, 352-355.

Bates, J. M. / Granger, C. W. J. (1969): The Combination of Forecasts, Operational Research Quarterly, 20, 451-468.

Beechey, M. / Österholm, P. (2014): Policy Interest-Rate Expectations in Sweden: A Forecast Evaluation, Applied Economics Letters, 21(13/15), 984-991.

Belongia, M. T. (1987): Predicting Interest Rates: A Comparison of Professional and Market-Based Forecasts, Federal Reserve Bank of St. Louis Review, 69(3), 9-15.

Benke, H. (2004): Kapitalmarktprognosen auf dem Prüfstand, Stiftung/Sponsoring, 4, 25-28.

Benke, H. (2006): Was leisten Kapitalmarktprognosen? – die Sicht eines Stiftungsmanagers, Zeitschrift für das gesamte Kreditwesen, 59(17), 902-906.

Brooks, R. / Gray, J. B. (2004): History of the Forecasters, Journal of Portfolio Management, 31(1), 113-117.

Butter, den F. A. G. / Jansen, P. W. (2013): Beating the Random Walk: A Performance Assessment of Long-Term Interest Rate Forecasts, Applied Financial Economics, 23, 749-765.

Cho, D. W. (1996): Forecasts Accuracy: Are Some Business Economists Consistently Better than Others? Business Economics, 31, 45-49.

Chortareas, G. / Jitmaneeroj, B. / Wood, A. (2012): Forecast Rationality and Monetary Policy Frameworks: Evidence from UK Interest Rate Forecasts, Journal of International Financial Markets, Institutions and Money, 22(1), 209-231.

Chun, A. L. (2009): Forecasting Interest Rates and Inflation: Blue Chip Clairvoyants or Econometrics?, EFA 2009 Bergen Meeting Paper, SSRN, 946667.

Diebold, F. X. / Lopez, J. A. (1996): Forecast Evaluation and Combination, Maddala, G. S. / Rao, C. R. (Ed.), Handbook of Statistics, 14, Amsterdam.

Diebold, F. X. / Mariano, R. S. (1995): Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, 253-263.

Domian, D. L. (1992): Money Market Mutual Fund Maturity and Interest Rates, Journal of Money, Credit and Banking, 24(4), 519-527.

Dua, P. (1988): Multiperiod Forecasts of Interest Rates, Journal of Business and Economic Statistics, 6(3), 381-384.

Francis, J. (1991): Management Anticipation of Interest Rates: The Case of Commercial Banks, Journal of Business Finance and Accounting, 18, 675-695.

Friedman, B. M. (1980): Survey Evidence on the 'Rationality' of Interest Rate Expectations, Journal of Monetary Economic, 6, 453-465.

Goodhart, C. A. E. / Lim, W. B. (2008): Interest Rate Forecasts: A Pathology, IJCB Discussion Paper, 612.

Gosnell, T. F. / Kolb, R. W. (1997): Accuracy of International Interest Rate Forecasts, Financial Review, 32(3), 431-448.

Granger, C. W. J. / Newbold, P. (1973): Some Comments on the Evaluation of Economic Forecasts, Applied Economics, 5(1), 35-47.

Greer, M. (2003): Directional Accuracy Tests of Long-Term Interest Rate Forecasts, International Journal of Forecasting, 19(2), 291-298.

Gubaydullina, Z. / Hein, O. / Spiwoks, M. (2011): The Status Quo Bias of Bond Market Analysts, Journal of Applied Finance and Banking, 1(1), 31-51.

Hafer, R. W. / Hein, S. E. (1989): Comparing Futures and Survey Forecasts of Near-Term Treasury Bill Rates, Federal Reserve Bank of St. Louis Review, 5, 33-42.

Hafer, R. W. / Hein, S. E. / MacDonald, S. S. (1992): Market and Survey Forecasts of the Three-Month Treasury-Bill Rate, Journal of Business, 65(1), 123-138.

Henriksson, R. D. / Merton, R. C. (1981): On Market Timing and Investment Performance, II. Statistical Procedures for Evaluating Forecasting Skills, Journal of Business, 54, 513-533.

Ilmanen, A. (1996): Market Rate Expectations and Forward Rates, Jounal of Fixed Income, 6(2), 8-22.

Jongen, R. / Verschoora, W. F. C. / Wolff, C. C. P. (2011): Time-Variation in Term Premia: International Survey-Based Evidence, Journal of International Money and Finance, 30, 605-622.

Joutz, F. / Stekler, H. O. (2000): An Evaluation of the Predictions of the Federal Reserve, International Journal of Forecasting, 16, 17-38.

Knüppel, M. / Schultefrankenfeld, G. (2013): The Empirical (Ir)relevance of the Interest Rate Assumption for Central Bank Forecasts, Deutsche Bundesbank Discussion Paper, 11/2013.

Kolb, R. A. / Stekler, H. O. (1996): How Well do Analysts Forecast Interest Rates?, Journal of Forecasting, 15(5), 385-394.

Kunze, F. / Gruppe, M. (2014): Performance of Survey Forecasts by Professional Analysts: Did the European Debt Crisis Make it Harder or Perhaps Even Easier? Social Sciences, 3, 128-139.

Kunze, F. / Gruppe, M. / Wendler, T. (2015): Interest Rate Forecasting and the Financial Crisis: A Turning Point in More than Just one Way, International Journal of Financial Engineering and Risk Management, 2(1), 1-16.

Kunze, F. / Kramer, J. / Rudschuck, N. (2014): Interest Rate Forecasts in Times of Financial Crisis: What Might be Interesting to Know? European Journal of Political Economy, 34, 45-52.

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 and Money, 48, 192-205.

Laplace, P. S. (1814): Essai Philisophique Sur les Probabilities, Paris.

Merton, R. C. (1981): On Market Timing and Investment Performance, I. An Equilibrium Theory of Value for Market Forecasts, Journal of Business, 54, 363-406.

Miah, F. / Khalifa, A. A. / Hammoudeh, S. (2016): Further Evidence on the Rationality of Interest Rate Expectations: A Comprehensive Study of Developed and Emerging Economies, Economic Modelling, 54, 574-590.

Mincer, J. / Zarnowitz, V. (1969): The Evaluation of Economic Forecasts, Mincer, J. (Ed.), Economic Forecasts and Expectation, Columbia University Press, New York, 3-46.

Mitchell, K. / Pearce, D. K. (2007): Professional Forecasts of Interest Rates and Exchange Rates: Evidence from the Wall Street Journal's Panel of Economists, Journal of Macroeconomics, 29(4), 840-854.

Mose, J. S. (2005): Expert Forecasts of Bond Yields and Exchange Rates, Danmarks National Bank, Monetary Review, 2005, 91-95.

Oliver, N. / Pasaogullari, M. (2015): Interest Rate Forecasts in Conventional and Unconventional Monetary Policy Periods, Federal Reserve Bank of Cleveland Economic Commentary, 5.

Pierdzioch, C. (2015): A Note on the Directional Accuracy of Interest-Rate Forecasts, Applied Economics Letters, 22(13), 1073-1077.

Scheier, J. / Spiwoks, M. (2006): Aktives Portfoliomanagement am britischen Anleihenmarkt, Meyer-Bullerdiek, F. / Spiwoks, M. (Ed.), Bank- und Finanzwirtschaft, 4, Frankfurt am Main.

Schwarzbach, C. / Kunze, F. / Rudschuck, N. / Windels, T. (2012): Asset Management in the German Insurance Industry: The Quality of Interest Rate Forecasts, Zeitschrift für die gesamte Versicherungswissenschaft, 101(5), 693-703.

Simon, D. P. (1989): The Rationality of Federal F/s Rate Expectations: Evidence from a Survey, Journal of Money, Credit, and Banking, 21, 388-393.

Spiwoks, M. (2003): Qualität der Zinsprognosen deutscher Banken – eine empirische Analyse, Kredit und Kapital, 36(3), 289-308.

Spiwoks, M. / Hein, O. (2007): Die Währungs-, Anleihen- und Aktienmarktprognosen des Zentrums für Europäische Wirtschaftsforschung – Eine empirische Untersuchung des

Prognoseerfolges von 1995 bis 2004, AStA-Wirtschafts- und Sozialstatistisches Archiv, 01, 43-52.

Spiwoks, M. / Bedke, N. / Hein, O. (2008): Forecasting the Past: The Case of US Interest Rate, Financial Markets and Portfolio Management, 22(4), 357-379.

Spiwoks, M. / Bedke, N. / Hein, O. (2009): The Pessimism of Swiss Bond Market Analysts and the Limits of the Sign Accuracy Test – An Empirical Investigation of their Forecasting Success Between 1998 and 2007, International Bulletin of Business Administration, 4, 6-19.

Spiwoks, M. / Bedke, N. / Hein, O. (2010): Topically Orientated Trend Adjustment and Auto-correlation of the Residuals – An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany, Journal of Money, Investment and Banking, 14, 16-35.

Spiwoks, M. / Gubaydullina, Z. / Hein, O. (2015): Trapped in the Here and Now – New Insights into Financial Market Analyst Behavior, Journal of Applied Finance and Banking, 5(1), 35-56.

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

Thiele, J. (1993): Kombination von Prognosen, Physica Verlag, Heidelberg.

Throop, A. W. (1981): Interest Rate Forecasts and Market Efficiency, Economic Review of the Federal Reserve Bank of San Francisco, Spring, 29-43.

Zarnowitz, V. / Braun, P. (1992): Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisions of Forecasting Performance, NBER Working Paper, 3965.

Appendix

Table 20: Overview of Studies on the Accuracy of Survey-Based Interest Rate Forecasts

Study	Countries Analyzed	Interest Rates Analyzed	Data Source	Period Considered	Methods Used	Result
Friedman (1980)	USA	Fed Funds Rate, 3-Month and 12-Month Bills, 6-Month Eurodollars, Utility Bonds, Municipal Bonds	Goldsmith-Nagan Bond and Money Market Letter	1969 - 1977	Unbiasedness Test, Efficiency Test, Consistency Test	Negative
Throop (1981)	USA	3-Month Treasury Bill Rate	Goldsmith-Nagan Bond and Money Market Letter	1970 - 1979	MSE, RMSE	Positive
Belongia (1987)	USA	3-Month Treasury Bill Rate	Wall Street Journal	1981 - 1986	Direction of Change, MAE, RMSE	Negative
Dua (1988)	USA	3-Month and 12-Month Treasury Bill Rate, Fed Funds Rate	Goldsmith-Nagan Bond and Money Market Letter / Federal Reserve Bulletin / Bond Buyer	1972 - 1985	MAE, RMSE, Theil´s U	Mixed
Simon (1989)	USA	Fed Funds Rate	Money Market Services	1984 - 1987	MAE, MSE	Negative
Hafer / Hein (1989)	USA	3-Month Treasury Bill Rate	Bond and Money Market Letter	1969 - 1989	Bias Tests, Market Efficiency Tests	Negative
Francis (1991)	USA	Various Bank Interest Rates in Pennsylvania	Call Reports	1983 - 1986	Mann-Whitney Test	Negative
Zarnowitz / Braun (1992)	USA	3-Month Treasury Bill Rate	ASA-NBER Quarterly Survey	1968 - 1990	ME, MAE, RMSE	Mixed
Hafer / Hein / MacDonald (1992)	USA	3-Month Treasury Bill Rate	Bond and Money Market Letter / Wall Street Journal	1977 - 1988	Unbiasedness Test, ME, MAE, RMSE, Theil´s U	Negative
Domian (1992)	USA	3-Month Treasury Bill Rate	IBC / Donoghue's Money Fund Report	1982 - 1990	Granger Causality	Negative
Ilmanen (1996)	USA	3-Month Treasury Bill Rate and 30-year Government Bond Yield	Wall Street Journal	1981 - 1994	Yield Change Predictions Compared to Forwards and No-Change	Negative

Study	Countries Analyzed	Interest Rates Analyzed	Data Source	Period Considered	Methods Used	Result
Kolb / Stekler (1996)	USA	3-Month Treasury Bill Rate and 30-Year Government Bond Yield	Wall Street Journal	1982 - 1990	Compared to No-Change, Random Walk measured by Skillings-Mack, Fisher's Exact	Negative
Cho (1996)	USA	3-Month Treasury Bill Rate and 30-Year Government Bond Yield	Wall Street Journal	1989 - 1994	Rank Consistency Test	Mixed
Gosnell / Kolb (1997)	GER, JPN, CH, GB, USA	3-Month Euromarket Rate	Risk	1990 - 1992	Measured against No-Change Model and Forward Rate Forecast	Mixed
Baghestani / Jung / Zuchegno (2000)	USA	3-Month Treasury Bill Rate	ASA-NBER Quarterly Survey	1983 - 1995	Unbiasedness Test	Negative
Albrecht (2000)	GER	3-Month Rate, 10-Year Government Bond Yield	Finanzen	1991 - 1997	ME	Negative
Spiwoks (2003)	GER	10-Year Government Bond Yield	Consensus Forecasts	1989 - 1999	Theil's U, TOTA Coefficient	Negative
Greer (2003)	USA	30-Year Government Bond Yield	Wall Street Journal	1984 - 1998	Binomial Test, Directional Accuracy Test, Institutional Affiliation Test	Mixed
Brooks / Gray (2004)	USA	30-Year and 10-Year Government Bond Yield	Wall Street Journal	1982 - 2002	Simplified Sign Accuracy Test, Simplified Unbiasedness Test	Negative
Benke (2004)	GER	10-Year Government Bond Yield	Handelsblatt	1991 - 2003	Simplified Sign Accuracy Test	Negative
Mose (2005)	GER, USA	10-Year Government Bond Yield	Consensus Forecasts	1989 - 2005	MAE	Negative
Baghestani (2005)	USA	3-Month Treasury Bill Rate	Survey of Professional Forecasters (SPF)	2001 - 2003	ME, MAE, RMSE	Negative
Scheier / Spiwoks (2006)	GB	10-Year Government Bond Yield	Consensus Forecasts	1989 - 2004	Theil's U_2, TOTA Coefficient	Mixed

Study	Countries Analyzed	Interest Rates Analyzed	Data Source	Period Considered	Methods Used	Result
Benke (2006)	GER	10-Year Government Bond Yield	Handelsblatt	1992 - 2005	Simplified Sign Accuracy Test	Negative
Spiwoks / Hein (2007)	FRA, GER, ITA, JPN, GB, USA	10-Year Government Bond Yield	ZEW-Finanzmarktreport	1995 - 2004	RMSE, MARE	Negative
Mitchell / Pearce (2007)	USA	3-Month Treasury Bill Rate and 30-Year Government Bond Yield	Wall Street Journal	1982 - 2002	Unbiasedness Test	Negative
Tabak / Feitosa (2008)	BRA	Short Term Interest Rate	Selic / Bloomberg and Central Bank of Brazil	1982 - 2002	MSE, Diebold-Mariano	Positive
Goodhart / Lim (2008)	NZ, GB	3-Month Official Cash NZ, Official Bank Rate UK (Libor)	RBNZ and BoE Interest Rate Forecasts	NZ 2000 - 2006 UK 1992 - 2004	Unbiasedness Test	Mixed
Spiwoks / Bedke / Hein (2008)	USA	10-Year Government Bond Yield and 3-Month Treasury Bill Rate	Consensus Forecasts	1989 - 2004	Unbiasedness Test, Sign Accuracy Test, Efficiency Test	Mixed
Spiwoks / Bedke / Hein (2009)	CH	3-Month Interest Rate and 10-Year Government Bond Yield	Consensus Forecasts	1998 - 2007	Unbiasedness Test, Sign Accuracy Test, TOTA Coefficient, Efficiency Test	Negative
Chun (2009)	USA	Fed Funds Rate, Short, Medium and Long Maturity Yield	Blue Chip Financial Forecasts	1993 - 2011	Compared against Time-series Models, Parametric Yield Curve Models and Futures Prices	Mixed
Spiwoks / Bedke / Hein (2010)	GER	3-Month Interest Rate and 10-Year Government Bond Yield	Consensus Forecasts	1989 - 2006	Unbiasedness Test, TOTA Coefficient, Efficiency Test, Sign Accuracy Test, Modified Diebold-Mariano Test, Theil's U_2	Mixed

Study	Countries Analyzed	Interest Rates Analyzed	Data Source	Period Considered	Methods Used	Result
Gubaydullina / Hein / Spiwoks (2011)	CAN, CH, ESP, FRA, GER, ITA, JPN, NLD, NOR, SWE, GB, USA	10-Year Government Bond Yield and 3-Month Interest Rate	Consensus Forecasts	1989 - 2009	TOTA Coefficient	Negative
Jongen / Verschoor / Wolff (2011)	23 countries inter alia AUS, HK, IDN, MYS, NZ, SGP, TWN	3-Month Interest Rates	Consensus Forecasts	1995 - 2009	Dickey-Fuller Unit Root Test, Expectations Hypothesis Tests	Mixed
Schwarzbach / Kunze / Rudschuck / Windels (2012)	GER	10-Year Government Bond Yield	Bloomberg, Reuters	1999 - 2011	Augmented Dickey Fuller Test (ADF Test), Johansen Approach, Granger Causality	Negative
Chortareas / Jitmaneeroj / Wood (2012)	GB	3-Month Interest Rate and 10-Year Government Bond Yield	Consensus Forecasts	1989 - 2006	Unbiasedness Test, Orthogonality Test	Negative
Baghestani / Marchon (2012)	BRA	Central Bank of Brazil Selic Interest Rate Target	Central Bank of Brazil Online Survey	2003 - 2011	Unbiasedness Test	Positive
Butter / Jansen (2013)	GER, GB, JPN, NLD, USA	10-Year Government Bond Yield	Consensus Forecasts	2003 - 2008	Successful Forecasts as a Percentage of Total Forecasts	Negative
Kunze / Kramer / Rudschuk (2013)	EUR	3-Month EURIBOR	Bloomberg/Reuters Professional Survey Forecasts	1998 - 2011	Granger Causality	Mixed
Knüppel / Schulte-frankenfeld (2013)	BRA, GB	Interest Rates Central Bank	COPOM, IBGE	1999 - 2011	RMSE	Positive
Kunze / Gruppe (2014)	EUR	3-Month EURIBOR	Consensus Forecasts	1998 - 2013	Quandt-Andrews Breakpoint Test, Theil´s U	Mixed
Baghestani / Danila (2014)	CZE	2-Week Repo Rate and 12-Month Interbank Interest Rate (PRIBOR)	Czech National Bank (CNB)	2005 - 2012	Theil´s U, Diebold-Mariano Test, Fisher´s Exact Test	Mixed

Study	Countries Analyzed	Interest Rates Analyzed	Data Source	Period Considered	Methods Used	Result
Beechay / Österholm (2014)	SWE	Government Bond Yield, Forward Rate and Interest-Rate Swaps	Prospera, Swedish Financial Markets	2002 - 2012	Unbiasedness Test, Efficiency Test, Modified Diebold Mariano Test, RMSE	Mixed
Kunze / Gruppe / Wendler (2015)	EUR	3-Month EURIBOR	Consensus Forecasts	1998 - 2013	Sign Accuracy Test, Turning Point Analysis, RMSE	Mixed
Spiwoks / Gubaydullina / Hein (2015)	CAN, CH, ESP, GER, FRA, GB, ITA, JPN, NLD, NOR, SWE, USA	10-Year Government Bond Yield	Consensus Forecasts	1989 - 2009	TOTA Coefficient	Negative
Oliver / Pasaogullari (2015)	USA	Fed Funds Rate, 1-Year, 5-Year and 10-Year Bond Yield	Blue Chip Financial Forecasts	1990 - 2012	RMSE	Mixed
Baghestani / Arzaghi / Kaya (2015)	AUS, CAN, CH, EUR, GB, JPN, USA	3-Month Eurocurrency Rate and 10-Year Government Bond Yield	Blue Chip Financial Forecasts	1999 - 2008	Unbiasedness Test, Theil's U, ME, MAE, Sign Accuracy Test, Rationality Test	Mixed
Pierdzioch (2015)	USA	Prime Interest Rate, Treasury Bill Rate and T-Bond Rate	Livingston Survey	1981 - 2013	ROC (Relative Operating Characteristic) Curves Analysing Techniques	Positive
Miah / Khalifa / Hammoudeh (2016)	30 Countries inter alia CHN, HK, IND, KOR, PHL, SGP, THAI, TWN	Long-Term and Short-Term Interest Rates	Fx4casts.com	2001 - 2012	Unbiasedness Test, Efficiency Test, Unit Root Test	Mixed
Kunze / Wegener / Bizer / Spiwoks (2017)	GER, GB	3-Month Interbank Rate and 10-Year Government Bond Yield	Consensus Forecasts	1993 - 2014	RMSE, Theil´s U, Diebold-Mariano Test	Negative

Declaration for Admission to the Doctoral Examination

I confirm

1. that this dissertation that I submitted 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 in-validation or revoking of the doctoral degree.