TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Improving Risk Management: Integrating and Forecasting Behavioral Aspects

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der Sozial- und Wirtschaftswissenschaften
unter der Leitung von

A.o. Prof. Mag. Dr. Wolfgang Aussenegg
E330 - Institut für Managementwissenschaften

eingereicht an der Technischen Universität Wien

Fakultät für Maschinenwesen und Betriebswissenschaften

von

Mag. Donald Baillie
Matrikelnummer 8650806
Porzellangasse 50/15, 1090 Wien

Wien, am 30. April 2012

Acknowledgements

Many people provided the inspiration necessary for an undertaking of the size and effort of this dissertation. First and foremost, I would like to thank Wolfgang Aussenegg at the Technical University and Erich Kirchler at the University of Vienna for graciously agreeing to support me in my endeavor and to supervise this thesis. Their input and feedback was invariably constructive and challenging. Both their views of the big picture as well as their attention to detail lifted the quality of research to a new level each time.

Without Thomas Oberlechner at Webster University this project would never have happened. His enthusiasm and support in researching the initial topics started me on the journey. His wonderfully positive feedback was always extremely welcome. My thanks also go to Dorothy and Michael Kopel for encouraging me to enquire at the Institute of Management Science at the Technical University. Peter Schaller of Bank Austria (and the Technical University) was also a vital friend in my quest to find initial support for the thesis. Gerhard Deschkan and Reinhard Keider from Bank Austria provided much-needed anonymized trading transactions for the tests on real data - many thanks for this vital support.

Many thanks also to Christian Kreuzer, then of the FHW University of Applied Science in Vienna. His support made the prediction market platform possible. Thanks also to Michael Palka of the HTL Mödling for graciously doing all he could to enable and initiate the research project at his school, nothing beats long-standing friendship.

Finally, my deepest gratitude goes to Magdalena Höbarth for being with me in every phase of this undertaking. Without her constant support and encouragement whenever the goal seemed too far away, I would never have managed to put in the extra work that always seemed necessary. Her support made it all worthwhile.

Donald Baillie
Vienna, April 2012

Abstract

Cognitive biases are systematic deviations from normative criteria that often lead to errors in judgment (Sander and Scherer, 2009). They are not new to psychologists - researchers in the fields of social cognition (see e.g. Fiske and Taylor, 2007) or attribution theory (see e.g. Heider, 1958) have identified that people are not always logical or rational when forming judgments about others (see Aronson et al., 2010). However, since Herbert Simon and Daniel Kahneman won the Nobel Memorial Prize in Economics in 1978 and 2002 respectively, research and public awareness of cognitive biases and the field of behavioral finance have expanded rapidly. The online platform Wikipedia (2011) currently lists more than 90 known cognitive biases. Of all these biases, Kahneman himself considers the effects of overconfidence to be the most catastrophic (Hubbard, 2009, p. 102).

This dissertation aims to find out whether the behavioral biases overconfidence and the related disposition effect can not only be measured, but also whether and how they can be forecast or whether early signs of these biases can be detected. An experimental market is used to generate data for analysis of behavioral biases. An online prediction market was set up on the European soccer championships 2008, and 340 participants generated almost 50,000 trades. The efficiency and prediction quality of the market is analyzed in order to research behavioral biases under experimental conditions. The efficiency of the market is established despite many of the participants being prone to behavioral biases.

Overconfidence on the experimental prediction market and the reasons for overconfidence and potential consequences in trading behavior are explored. Two measures for detecting overconfidence are developed, an overconfidence index (OCI), and a degree of diversification (DoD) measure. Further, potential connections between overconfidence and trading strategies are investigated, and overconfidence (and conversely, underconfidence) is found to be related to character and to be prevalent throughout the investment experiment.

This forms the basis for finding ways of pre-detecting overconfidence. Statistical procedures such as multiple regression, factor analysis and correpondence analysis are used for this task. Suggestions are made on how to predict potential overconfidence when trading or investing for risk management purposes.

Analysis is performed to test for the disposition effect on data from the prediction market as well as on data from a real foreign exchange (FX) market. It is shown that the disposition effect is related to overconfidence on both the experimental market and the real FX market. This research finds that even though the disposition effect is larger on the experimental market made up of investment amateurs, it is still present for professionals on the real FX market, albeit in lower intensity and in different strength for different portfolios.

Tests for overconfidence outside the financial world are performed. The results show that recognition and management of behavioral biases is not restricted to finance, rather tests for overconfidence in engineering students and professionals show that behavioral biases are just as prevalent in a technical field as in finance.

Finally, the risk management process is discussed and an indication of where the insights from the previous chapters can be applied to risk management is provided. The results on overconfidence and the disposition effect lead to the insight that when setting up risk management processes care should be taken to devise individualized approaches rather than generally devising limiting strategies of the one-size-fits-all type.

It can be derived from the analysis of these cognitive biases that any systems and processes for prediction and early warning of losses due to behavioral biases must be structured differently from existing risk management processes. Information for detection of behavioral biases must either be of a longitudinal nature, i.e. comparisons of ex-ante data with ex-post realizations in the case of the overconfidence bias, or real-time analyses and monitoring in the case of the disposition effect.

Kurzzusammenfassung

Kognitive Verzerrungen (cognitive biases) sind systematische Abweichungen von normativen Kriterien die oft zu Fehlern bei Beurteilungen oder Entscheidungen führen (Sander and Scherer, 2009). Sie sind für Psychologen nicht neu - Forscher in den Bereichen der sozialen Kognition (vgl. z.B. Fiske and Taylor, 2007) oder der Attributionstheorie (vgl. z.B. Heider, 1958) haben identifiziert, dass Personen nicht immer logisch oder rational handeln wenn sie Urteile oder Entscheidungen in Bezug auf andere Personen formulieren (vgl. Aronson et al., 2010). Seitdem Herbert Simon 1978 und Daniel Kahneman 2002 den Wirtschaftsnobelpreis (korrekt: der von der schwedischen Reichsbank in Erinnerung an Alfred Nobel gestiftete Preis für Wirtschaftswissenschaften) bekamen, hat die Forschung sowie die öffentliche Wahrnehmung von kognitiven Verzerrungen und dem Gebiet der verhaltensorientierten Finanzierungslehre (behavioral finance) rapide zugenommen. Die Onlineplattform Wikipedia (2011) listet gegenwärtig mehr als 90 bekannte kognitive Verzerrungen auf. Von allen diesen Verzerrungen befindet Kahneman selber, dass die Effekte der Selbstüberschätzung (overconfidence) am Schwersten wiegen (Hubbard, 2009, S. 102).

Diese Dissertation hat das Ziel herauszufinden, ob Selbstüberschätzung als Verhaltensneigung und der damit verwandte Dispositionseffekt (disposition effect) nicht nur gemessen, sondern auch vorhergesagt werden können, sowie ob Frühwarnsignale dieser Neigungen erkannt werden können. Ein Experimentalmarkt wird hier verwendet um Daten für die Analyse der Verhaltensverzerrungen zu generieren. Es wird ein Online Vorhersagemarkt (prediction market) auf die Fussballeuropameisterschaft 2008 aufgesetzt. Daraus generieren 340 Teilnehmer fast 50,000 Transaktionen. Die Effizienz und die Vorhersagequalität werden analysiert, um Verhaltensverzerrungen unter experimentellen Bedingungen zu erforschen. Die Effizienz dieses Marktes kann bestätigt werden, obwohl viele der Teilnehmer Verhaltensverzerrungen aufweisen.

Mögliche Selbstüberschätzung der Teilnehmer auf dem Experimentalmarkt, sowie die Gründe für Selbstüberschätzung und potentielle Konsquenzen davon im Handelsverhalten werden erforscht. Zwei Masszahlen für die Erkennung von Selbstüberschätzung werden entwickelt - ein Selbstüberschätzungsindex (overconfidence index, OCI), sowie

eine Masszahl der Portfoliodiversifikation (degree of diversification, DoD). Weiters werden mögliche Verbindungen zwischen Selbstüberschätzung und Handlesstrategien untersucht. Selbstüberschätzung (sowie auch Selbstunterschätzung) kann mit Charaktereigenschaften in Verbindung gebracht werden und ist während des gesamten Investitionsexperiments zu beobachten.

Dies bildet die Basis für Methoden um Selbstüberschätzung bereits ex-ante, vor Eintritt der Realisation, zu erkennen. Für die Analyse der Zusammenhänge zwischen ex-ante und Realisation von Selbstüberschätzung werden die statistische Methoden multiple Regression, Faktorenanalyse und Korrespondenzanalyse eingesetzt. Darauf aufbauend werden Vorschläge für die Vorhersage von möglicher Selbstüberschätzung im Handel oder Investment für Risikomanagementzwecke erarbeitet.

Weiters werden in der vorliegenden Arbeit Tests zur Messung des Dispositionseffekts, auf Basis von Daten aus dem Vorhersagemarkt sowie aus Echtdaten auf einem realen Wechselkurs (FX) Markt, durchgeführt. Es wird gezeigt, dass der Dispositionseffekt sowohl auf dem Experimentalmarkt wie auch auf dem echten FX Markt mit Selbstüberschätzung verbunden ist. Die vorliegenden Untersuchungen zeigen, dass der Dispositionseffekt sowohl bei Amateurinvestoren am Experimentalmarkt, als auch bei professionellen Händlern am realen FX Markt zu beobachten ist. Der Effekt ist bei den Amateuren stärker zu beobachten als bei den professionellen Händlern, bei denen der Effekt pro Portfolio unterschiedlich stark ausfällt.

Zusätzlich werden Tests für Selbstüberschätzung ausserhalb der Finanzwelt durchgeführt. Die Resultate zeigen, dass die Erkennung und das Management von Verhaltensverzerrungen nicht auf die Finanzwelt beschränkt ist. Tests auf Selbstüberschätzung bei Technikstudenten und professionellen Zivilingenieuren zeigen, dass Verhaltensverzerrungen in der Technik ebenso wie in der Finanzwelt präsent sind.

Schliesslich wird der Risikomanagementprozess dargestellt und eine Indikation gegeben, wo die Erkenntnisse der vorhergehenden Kapitel im Risikomanagement angewendet werden können. Die Resultate der Untersuchungen bezüglich Selbstüberschätzung und dem Dispositionseffekt führen zu der Erkenntnis, dass beim Aufsetzen von Risikomanagementprozessen Acht gegeben werden muss, dass individualisierte Ansätze zur Limitierung von Transaktionen implementiert werden, anstatt wie bisher meist auf generellen Einheitswerten basierende Ansätze.

Es kann aus der Analyse dieser kognitiven Verzerrungen abgeleitet werden, dass Systeme und Prozesse für die Vorhersage und für Frühwarnindikationen von Verlusten aus verhaltensorientierten Verzerrungen anders aufgebaut werden müssen als in der bisherigen Praxis der Risikomanagementprozesse. Informationen zur Erkennung von Verhaltensbiases müssen entweder aus Längsschnittdaten bestehen, das heisst aus dem Vergleich von ex-ante Daten mit ex-post Realisationen bei der Untersuchung von Selbstüberschätzung, oder aus Echtzeitdaten mit begleitendem Monitoring bei der Erkennung des Dispositionseffektes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

On March 11, 2011, the BBC reported that "*Japan's most powerful earthquake since records began has struck the north-east coast, triggering a massive tsunami. [...] A state of emergency has been declared at a nuclear power plant, where pressure has exceeded normal levels*" (BBC, 2011b). On March 12, the BBC reports that "*[a]n evacuation zone around the damaged nuclear plant has been extended to 20km (12.4 miles) from 10km, and a state of emergency declared. An estimated 200,000 people have been evacuated from the area*", and that "*[a] building housing a reactor was destroyed, but authorities said the reactor itself was intact. The government sought to play down fears of a meltdown at the Fukushima 1 plant*" (BBC, 2011a).

Even though risk management processes were in place and in use, the Washington Post reports on May 27, 2011, that "*[n]uclear fuel at the stricken Fukushima Daiichi power plant began melting just five hours after Japan's March 11 earthquake*" (Vastag and Mufson, 2011), with potentially disastrous long-term effects for the population and the environment.

This formed a major risk management failure on the part of the power company TEPCO, who ignored many warning signs. The Japanese newspaper, The Asahi Shimbun, reports in an article titled "*Nuclear expert: Hubris led to disaster*", that "*[...] a former laboratory chief at what is now the Japan Atomic Energy Agency, suggests that today's generations of scientists and plant operators may have been blinded by an over-confidence in Japanese technology, saying they must try to prepare for the unforeseeable*

(Sugiura and Sasaki, 2011).

Turning from engineering risk management to financial risk management, the latest scandal in a string of huge trading losses within the past few years occurred at the Swiss bank UBS. A September 16, 2011, cbcnews report ran an article titled "*UBS case latest failure of risk management*" (cbcnews, 2011). The article reports that "*[p]olice in London arrested a 31-year-old UBS AG employee early Thursday morning following allegations the giant Swiss bank was hit with unauthorized trading losses of US$2-billion in a case that bares striking similarities to a string of trading scandals elsewhere"* (cbcnews, 2011).

The newspaper continues with the opinion that "*The financial industry has taken major steps in the wake of such rogue trading incidents to prevent them from recurring, beefing up risk management and providing better oversight. However, the frequency and size of losses has changed little"* (cbcnews, 2011).

What could be the reason that even the newspapers (if not only financial institutions' internal risk managers) are of the opinion that risk management in the financial industry has been increased but is not seeming to function to the extent that would be necessary?

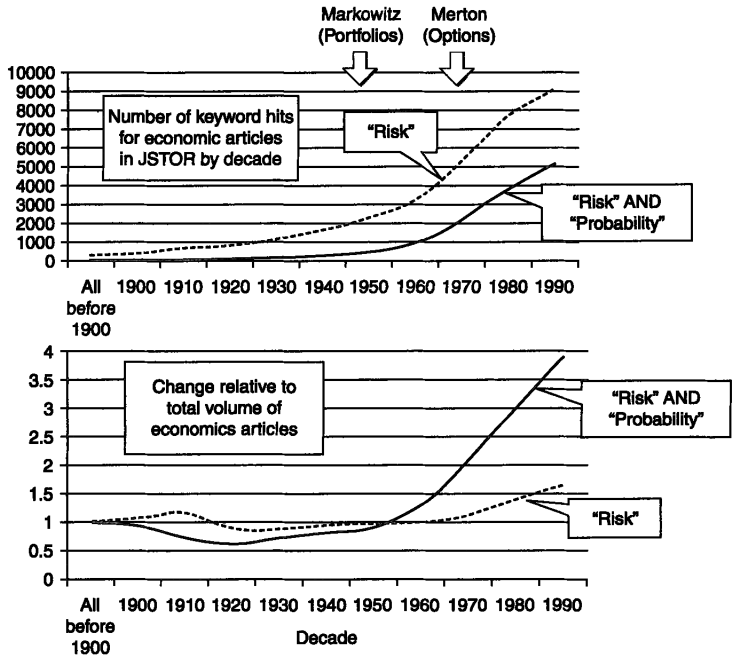
The Swiss Banking Commission finds that "*[i]n retrospect, insufficient attention to the inherent risk related to the balance sheet growth and the over-confidence in the existing risk management and risk control mechanisms appear as significant failures on the part of the bank*" (Commission, 2008).

Two topics are common to the examples presented above: First, the failure of risk management is a front page news item, both in an engineering context and in a finance context. Risk management has obviously entered the public domain. Second, the psychological aspect also features prominently; the word "overconfidence" appears repeatedly as an explanation for the outlined risk management failures.

This leads to the question of what is being done to prevent such risk management failures, if even the general public is aware of the connection between psychological terms such as the behavioral bias of overconfidence is somehow connected to risk management failures?

In recent years, attention in this area has focused on two aspects: Quantification of risks, and regulation of institutions. Since Markowitz (1952) and Samuelson (1938) led the development of mathematical and statistical models in finance and economics, these areas have become predominantly reliant on quantification. Hubbard (2009) researches academic databases for the occurrence of the words "risk" and "probability" together, in order to find out the development of the quantification of risk management since before the year 1900, shown in Figure 1.1. The increase in academic articles combining risk and probability analysis is dramatic.

Figure 1.1: Risk and Probability in Literature



Source: Hubbard, 2009, p. 65

However, a Zurich Report in Applied Risk Management cautions that "*Quantitative tools are important, but informed qualitative judgments are indispensable*" (Hofmann, 2008) and warns against an over-reliance on quantitative models to the exclusion of the understanding of the business and the involved persons.

Regulation of financial institutions has been a major point of attack to limit crashes and systemic problems by increasing required capital and focusing more attention on risk management, as documented by U.S. legislation such as the Sarbanes-Oxley Act (U.S. Congress, 2002), or in Europe by the Basel II (on Banking Supervision, 2006) and upcoming Basel III guidelines (on Banking Supervision, 2011).

Quantification and regulation are both attempting to reign in the excesses in part generated due to the ever-increasing complexity of trading and investing, but the human interaction component is always present, as evidenced in the quotes above dealing with (human) overconfidence as a corollary to financial or environmental disasters.

The field of finance has started to incorporate findings from psychology dealing with cognitive biases, mainly since Simon (1955) and Tversky and Kahneman (1974), resulting in a new field called behavioral finance (see Mullainathan and Thaler, 2000). Behavioral finance has come a long way since the seminal articles quoted above, but has not yet found its way into risk management within academia. This dissertation addresses this gap in the literature and finds new ways of integrating psychological analysis of cognitive biases into risk management.

1.2 Aims and Contributions

1.2.1 Aims

Cognitive biases are systematic deviations from normative criteria that often lead to errors in judgment (Sander and Scherer, 2009). They are not new to psychologists - researchers in the fields of social cognition (see e.g. Fiske and Taylor, 2007) or attribution theory (see e.g. Heider, 1958) have identified that people are not always logical or rational when forming judgments about others (see Aronson et al., 2010). However, since Herbert Simon and Daniel Kahneman won the Nobel Memorial Prize in Economics in 1978 and 2002 respectively, research and public awareness of cognitive biases and the field of behavioral finance have expanded rapidly. The online platform Wikipedia (2011) currently lists more than 90 known cognitive biases. Of all these biases, Kahneman himself considers the effects of overconfidence to be the most catastrophic (Hubbard, 2009, p. 102). Research (e.g. Odean, 1999) shows that investors who are prone to the effects of overconfidence also fall prey to the disposition effect (the propensity to sell winning investments too early and to hold on to losing investments for too long). Siwar (2011) finds that the overconfidence bias and the disposition effect are linked in many cases.

This dissertation aims to find out whether the behavioral biases overconfidence and the disposition effect can not only be measured, but also whether they can be forecast or whether early signs of these biases can be detected. If so, this research aims further to provide indications on how to set up an early warning system for risk managers, which types of information are necessary for the task, and which measures will have to be implemented. Within the dissertation, each chapter provides an interconnected but independent contribution toward this aim.

In **chapter 2**, an experimental market is used to generate data for analysis of behavioral biases. An online prediction market was set up on the European soccer championships 2008, and 340 participants generated almost 50,000 trades. The efficiency and prediction quality of the market is analyzed in order to research behavioral biases under experimental conditions. The efficiency of the market is established despite many of the participants being prone to behavioral biases.

Chapter 3 provides the details of the analysis of overconfidence on the experimental prediction market and the reasons for overconfidence and potential consequences in trading behavior are explored. In this chapter, two measures for detecting overconfidence are developed, an overconfidence index (OCI), and a degree of diversification (DoD) measure. The results of previous research such as by Odean (1998b, 1999) are presented in context and contrasted with this paper's results. They find trading frequency related to overconfidence, and that men trade more than women, amongst other results; but these results cannot be confirmed in the current data. Further, potential connections between overconfidence and trading strategies are investigated, and overconfidence (and conversely, underconfidence) is found to be related to character and to be prevalent throughout the investment experiment. This forms the basis for finding methods to pre-detect overconfidence.

In **chapter 4**, the methods of multiple regression, factor analysis and correpondence analysis are applied to the data from the prediction market to identify connections leading to whether and how one can predict overconfidence. These connections can be identified and suggestions are made on how to predict potential overconfidence when trading or investing for risk management purposes.

Analysis is performed in **chapter 5** to test for the disposition effect on data from the prediction market as well as on data from a real foreign exchange (FX) market. It is shown that the disposition effect is related to overconfidence on both the experimental market and the real FX market. Previous literature (e.g. by Odean, 1998a) or (Weber and Camerer, 1998) find the disposition effect on experimental as well as on real markets. Some authors such as O'Connell and Teo (2009) assert that professional traders are much less prone to the disposition effect than amateurs or retail investors, but authors such as Garvey and Murphy (2004) find ample evidence of the disposition effect amongs professionals.

This research finds that even though the disposition effect is larger on the experimental market made up of investment amateurs, it is still present for professionals on the real FX market, albeit in lower intensity and in different strength for different portfolios.

As the previous chapters deal with behavioral biases in finance, **chapter 6** provides the results of tests for overconfidence outside the financial world. These results show that recognition and management of behavioral biases is not restricted to finance, rather tests for overconfidence in engineering students and professionals show that behavioral biases are just as prevalent in a technical field as in finance. Methods for recognizing and managing behavioral biases in this context are just as important as in finance, even though the magnitude of risk management failures is not usually so newsworthy as when failings in finance are uncovered.

Chapter 7 deals with the risk management process and provides an indication of where the insights from the previous chapters can be applied to risk management. The results on overconfidence and the disposition effect lead to the insight that when setting up risk management processes care should be taken to devise individualized approaches rather than generally devising limiting strategies of the one-size-fits-all type.

It can be derived from the analysis of these cognitive biases that any systems and processes for prediction and early warning of losses due to behavioral biases must be structured differently from existing risk management processes. Information for detection of behavioral biases must either be of a longitudinal nature, i.e. comparisons of ex-ante data with ex-post realizations in the case of the overconfidence bias, or real-time analyses and monitoring in the case of the disposition effect.

1.2.2 Contributions

This dissertation makes the following contributions to the existing literature:

- A prediction market is established as a viable proxy for real markets, showing aggregate efficiency despite evidence of behavioral biases of participants on the market.
- A new measure for overconfidence, an overconfidence index, is developed.
- A new measure for portfolio diversification is developed, the degree of diversification.
- It is empirically shown that overconfidence can be predicted with analytic methods.
- It is confirmed that amateur investors on an experimental market are more prone to the disposition effect than professional traders in a real market.

- It is shown that the disposition effect occurs selectively amongst professional foreign exchange traders.
- The necessary information and analysis is shown for predicting and monitoring the behavioral biases overconfidence and the disposition effect.

The research in this dissertation is intended for publication in peer-reviewed journals and for presentation at conferences. The results from chapter 3 (overconfidence on a soccer prediction market) were the basis for a paper that was submitted to the Journal of Behavioral Finance, a paper containing the results from chapter 4 (forecasting overconfidence) is in preparation for submission to a journal and is also in preparation for entry to one of the major international conferences on behavioral economics and economic psychology. A paper based on the overconfidence data from chapter 3 has been accepted for presentation at the International Congress of Psychology 2012.

Chapter 2

A Soccer Prediction Market

2.1 Introduction

In his well-known book "Management", Peter Drucker (1974) finds of forecasting that *"... any attempt to do so is foolish; the future is unpredictable. We can only discredit what we are doing by attempting it."*, and that *"... forecasting is not a respectable human activity and not worthwhile beyond the shortest of periods."*. However, Drucker was wanting to make a point about the necessity of strategic planning rather than just equating forecasting with soothsaying. He asks the question *"What do we have to do today to be ready for an uncertain tomorrow?"*. Drucker was concerned with planning, but planning is often confused with forecasting.

The Oxford English Dictionary defines forecasting as *"to predict or estimate (a future event or trend)"* and prediction as *"to say or estimate that (a specified thing) will happen in the future or will be a consequence of something"*(Dictionaries, 2010).

Planning, on the other hand, is defined by the Dictionary as *"to decide on and make arrangements for in advance"*, which is a different activity: the act of decision-making is already implicit in the concept of planning. Forecasting or prediction takes place before any decision is taken and provided the basis for this decision. According to Armstrong (2002), planning concerns what the world should look like, while forecasting is about what it will look like.

So why forecast at all? On the one hand, an often-heard paraphrase of George Santayana' original quote states that *"Those who cannot learn from history are doomed* *to repeat it."* (For the original, see Santayana, 1905)[1]. Obviously, some form of information is necessary in order to formulate decisions, one method entailing taking past events as a basis for current decisions, indeed, when performing portfolio optimization under Markowitz (1952) Modern Portfolio Theory or Sharpe (1964) CAPM conditions, the optimal forecast for the portfolio return is the conditional expected value (based on past return data).

Without estimations of what the future could possibly entail, all decisions reduce to decisions under ignorance rather than decisions under risk. Since it was first formulated by Frank Knight (1921), the former has become generally recognized as "Knightian Uncertainty", separating those decisions in which the potential outcomes are known, but not which one will occur, from those decisions where even the set of outcomes is unknown. If future events are certain or directly controllable, no forecasting is necessary. If, however, there is uncertainty about the future, decision makers require methods to base their decisions upon.

2.2 Forecasting

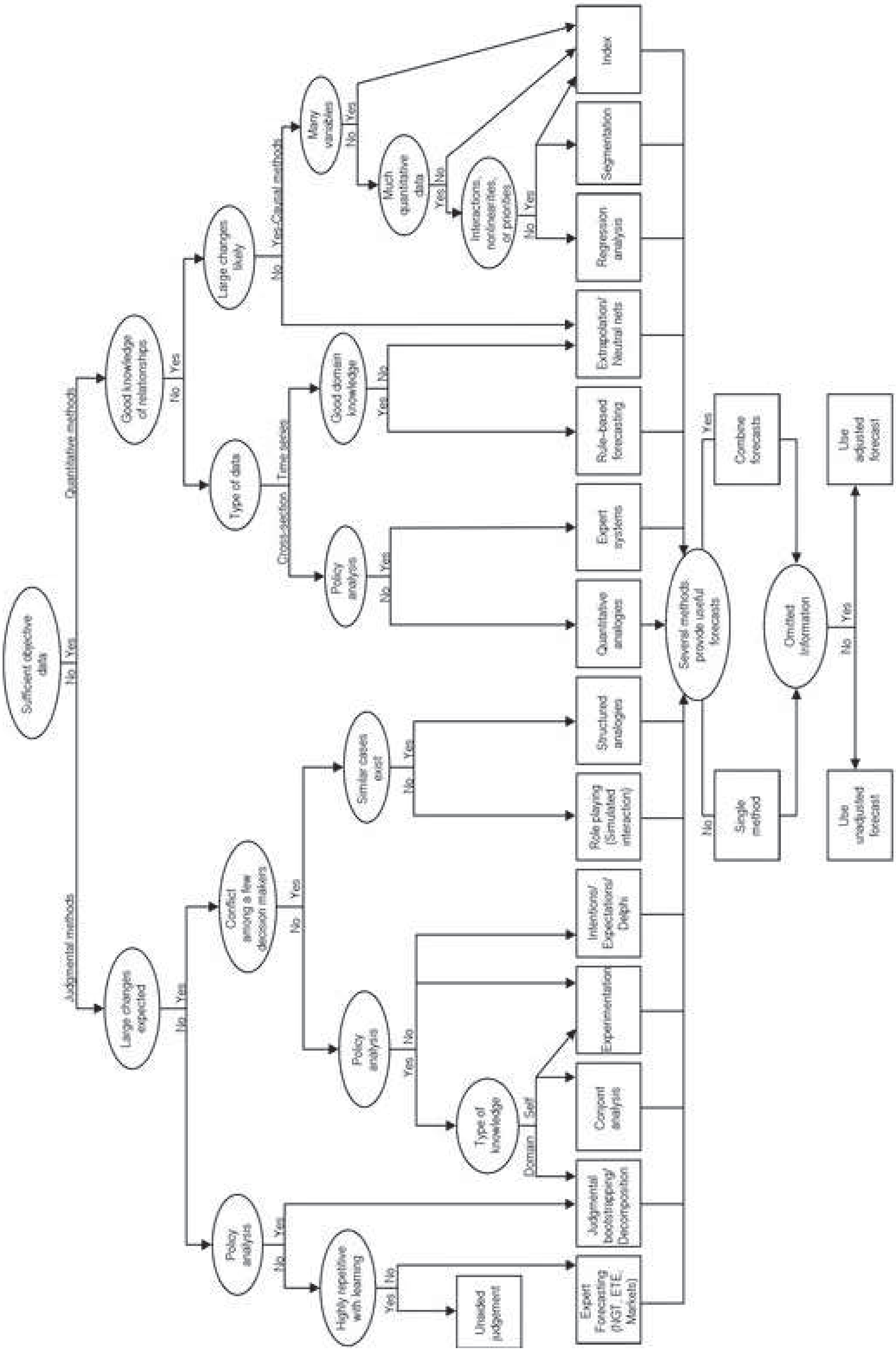
A comprehensive overview of forecasting methods is presented by Armstrong (2002), see Figure 2.1.

He separates forecasting methods into two distinct schools: methods that rely on the judgment of individuals, and quantitative methods. Quantitative methods include statistical or econometric methods such as regression analysis and expert systems, to mention only a few. The branch containing judgmental methods includes different individual or group methods such as expert forecasting, focus groups or the Delphi method. The other judgmental methodology involves prediction markets, which differs substantially from the former methods. Methods of forecasting that involve polling of individuals, collect information from group members individually and use simple statistical methods to aggregate the individual opinions without usually weighting opinions dependent upon levels of expertise or the such. Group deliberation methods bring individuals together and enable aggregation of opinions through discussion. One of the main drawbacks consists of social pressures or biases that cloud the objectivity of results.

Forecasts via opinions by individual experts are valuable techniques used in many

[1] The original quote is *Those who cannot remember the past are condemned to repeat it.*

Figure 2.1: Methods of Forecasting



Source: Armstrong, Principles of Forecasting, 2002, p. 376.

forms, but also suffer from drawbacks. In the investing arena, University professors such as Burton Malkiel have little time for the so-called expert investors. Burton Malkiel throws down the challenge that *"... a blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by experts"*. (Malkiel, 2007).

Group decisions have many advantages. In a study published by *Nature*, Giles (2005) finds that group efforts can at least match, if not outperform, expert opinions by comparing the online encyclopedia Wikipedia with the Encyclopaedia Britannica. He finds that the quality of Wikipedia, which is open to all contributors, is comparable to that of the Encyclopaedia Britannica, which relies on experts for its contributions.

Surowiecki (2004) quotes the famous example from Galton (1907) where the average of the crowd's estimate of a bull's weight is almost precisely the correct value. He also found that on the TV game show *"Who Wants to Be a Millionaire*, the expert was right 65% of the time, but the studio audience was right 91% of the time. This result is due to Condorcet's Jury Theorem (see Mueller, 2003): if the guesses of most of the members in an audience are better than random, and the other members' guesses are distributed randomly, then a plurality vote will ensure that the vote converges to the correct answer[2].

Group opinion aggregation is not without its pitfalls, however; problems with the aggregation of information have occupied economists for a long time. As mentioned above, expert opinions do not always guarantee the correct aggregation of information, and often succumb to biases such as groupthink, when members of a decision-making group prefer the desire for harmony to conflict versus their peers or hierarchical superiors. The term was first introduced to describe this phenomenon by Whyte in 1972 (see e.g. Whyte, 1989; Janis, 1972). Systematic errors can occur when groups amplify individuals' cognitive biases. Biases that occur during group aggregation of individual judgments also include information cascading (Asch, 1955), when members of a group tend to conform to an established group norm rather than form their own conflicting opinion, sometimes despite overwhelming evidence that the group opinion is completely wrong. Information that is held by most or all group members is also likely to influence individual judgments, and information held by only a few members is often left unarticulated - leading to "hidden profiles" (see e.g. Stasser and Titus, 2003). Russo and Schoemaker (1989) provide an overview of ten of the most popular barriers to decision-making.

[2]If however, less than 51% of individuals vote the right answer, the probability converges to 0% rather than 100%.

Choosing an appropriate forecasting method depends on the situation. For example, for long-range forecasting of the environment or of the market, econometric methods are often appropriate. For short-range forecasting of market share, extrapolation methods are useful. Forecasts of new-product sales could be made judgmentally by experts. Decisions by parties in conflict, such as companies and their competitors, can be predicted by role-playing (Armstrong, 2002), see Figure 2.1.

Cowles (1933, 1944) concluded that forecasters could not improve the accuracy of forecasts derived from the actions of a market. Research findings since then have strengthened this conclusion (Sherden, 1998). This applies to financial markets as well as betting on sporting events. Using markets to aggregate opinions enable individuals to utilize public and private information and forces them to come to an aggregate opinion via the consensus price. Those market participants who disagree with the level of the price can bet against it by going short or selling the assets. Forecasting via markets is often performed via experimental markets such as prediction markets.

2.2.1 Experimental Markets

There has been much effort directed into attempting to align classical finance theory with how researchers (in the beginning mainly psychologists) observed human judgment and decision making to function in practice. Researching behavior of individuals on markets can take two forms: observation or experiment. In a financial sense, either the researcher can observe behavior "in the wild" on existing markets, or create a market expressly for the purpose of observing the behavior patterns of interest by experiment.

Experimental markets have not often been used to test economic questions compared to theoretical or empirical analysis, as experimentation was seen for a long time as inferior to establishing theoretical models and testing against them (Bardsley et al., 2010). Vernon Smith has been a vociferous supporter of using experiments in economics and finance analogously to usage in the natural sciences. *"Economics has tended to be long on theories (Hypotheses) and the use of "logical completeness" as a criterion for judging the value of a theory, but short on technologies for discriminating among theories on the basis of rigorous standards of empirical evidence"* (Smith, 1982).

Following on from general (competitive) equilibrium theory founded by Walras (1874) and expanded by Arrow and Debreu (1954), free markets lead to an optimal Pareto-efficient allocation of resources and to efficiency in perfect markets, but under the condition that every buyer and seller has complete information. In a Cournot-Nash equilibrium established by Nash (1951) on the basis of Cournot's work (see Cournot, 1838), all individuals know what all other buyers and sellers are willing to pay or sell for and everyone knows that they know, i.e. all are perfectly rational by maximizing their self-interest. All agents are presumed to be price-takers in a Cournot equilibrium, as the number of buyers and sellers is so large that each individual has an imperceptible influence on price and, consequently, takes price as a given constant. Samuelson (1957) states it even more stringently in that competitive allocations require perfectly "foreseen" conditions of supply and demand.

Adam Smith suggests that the attainment of allocations under competitive equilibrium do not require any individual participant to have knowledge of the circumstances of other agents, or to have an understanding either of the market as an allocation system or of his/her role in promoting *"an end which was no part of his intention"* (Smith, 1776), and Alfred Marshall notes in his description of price determination in a local corn market, that *"it is not indeed necessary for our argument that any dealers should have a thorough knowledge of the circumstances of the market"* (Marshall, 1890). Hayek (1945) stated the case more strongly by emphasizing that *"the most significant fact about this (price) system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action ... "*.

Vernon Smith (1962) asks how "little" can the knowledge of each individual be, and still allow the market to efficiently allocation resources. The extreme case of "little" knowledge is the circumstance of strict privacy where each buyer and seller in a market knows only their own valuation of a commodity. He uses experimental markets to test what he calls the Hayek Hypothesis: *"Strict privacy together with the trading rules of a market institution are sufficient to produce competitive market outcomes at or near 100% efficiency"* (Smith, 1962).

Vernon Smith has commented at length on the over-reliance of economists on theories rather than on evidence-based methodologies. In 1948, Edward Chamberlin was the first to document an economic experiment when he set up a market under laboratory conditions to analyze whether the market was able to aggregate supply and demand information (Chamberlin, 1948). Thurston (1931) uses an experiment to derive a "satisfaction curve" expressing the fact that the satisfaction derived from any commodity is proportional to the logarithm of the quantity of the commodity. From this he derived the "indifference curve" showing the respective amounts of two commodities and the equivalent satisfaction level.

Systematic research into experimental markets under controlled conditions started at the end of the 1950's with Hoggatt (1959), Siegel and Fouraker (1960) or Smith (1962). In recent years, experimental economics have become much more popular, and today many researchers build on widely-accepted experiments by Reinhard Selten, Charles Plott and Vernon Smith. The fields of experimental economics and experimental psychology have become much more interwoven since Kahneman and Tversky's landmark "crossover" experiments (Kagel, 1995).

2.2.2 Prediction Markets

The Motivation for Online Markets

According to Wilson (1992), auction markets are an ideal vehicle for conducting experimental studies, as the explicitly fixed trading rules hold the rules of the game constant. Additionally, auctions play an important role in exchange theory, as they are one of the simplest methods of price determination without intermediating market makers, and behavior can be modeled practicably using auctions. In a double auction, buyers and sellers submit bids into an order book into two queues (the bid queue and the offer queue). Supply and demand are ranked according to the prices bid or offered, and matching occurs where supply and demand meet. A spread always exists between the highest bid and the lowest offer prices; the more liquid the market, the tighter the spread (for an overview of double auction markets, see e.g. Friedman and Rust, 1993). Continuous double auction markets allows trading at any time, with buyers and sellers free to continuously update their bids and offers (Kagel, 1995). Ever since the University of Iowa began its online Iowa Electronic Markets in 1988 to forecast the result of U.S. national and regional elections (see www.biz.uiowa.edu/iem), prediction markets have formed a very successful, though still not widely used, new branch of forecasting.

In an electronic prediction market, traders buy and sell assets, mostly in the form of a continuous double auction, whose final value is connected to some future outcome, i.e. election results, sporting events or the success of business applications. Tradable contracts are often set up as a futures market, the contracts being Arrow-Debreu (Arrow and Debreu, 1954) securities spanning the possible outcomes (e.g. election or sports competition results). Traders make a market in shares representing parimutuel claims on some outcome such as the outcome of an election. In parimutuel markets, the settlements occur at a well-defined end state known to all participants (e.g. such as after the election results have been published). Unlike in the stock markets, where expectations can fluctuate continuously with no defined endpoint, market odds on prediction markets approximate winning probabilities remarkably well (Thaler and Ziemba, 1988).

Prediction markets effectively offer the trader a binary option on the outcome of the event. Wolfers and Zitzewitz (2006) argue that the price of the traded security on a prediction market represents the probability of the event occurring. According to Ray (2006), the price of the claim in a prediction market is the market's consensus probability that an event will occur.

The idea is that, as outlined above, the price mechanism will focus all available public information as well as traders' private information into the price of the given asset. On a prediction market, the price serves to reveal all the relevant information that is in all of the traders' possession (Sunder, 1995). This implies that, via the price mechanism, all participating traders have access to all available information (Plott, 2000). Parimutuel betting markets such as prediction markets do have surprisingly efficient outcomes which are not attributable to individuals with large amounts of expertise, traders are rather "naive in the economic sense" (Smith, 2008). Smith also reports that despite the efficiency of such markets, there are still observable inefficiencies such as the "favorite-long-shot bias" (individuals tend to overvalue small probabilities and undervalue near certainties), but true to the efficient market hypothesis, Thaler and Ziemba find that arbitrageurs are able to take advantage of these inefficiencies and thus eliminating them.

The aggregation of public and private information from many sources into one price was made popular by Surowiecki (2004) in his best-selling book, "The Wisdom of Crowds", in which he explores the idea that large groups of people come to better decisions and predictions than individuals do by themselves. This effect of the price mechanism on markets goes all the way back to Hayek (1945), who first put forward the theory that prices aggregate all the frequently contradictory, incomplete and dispersed information held by a group of individuals. Smith (1982), in his well-known Hayek hypothesis, asserts that even when traders know very little about their environment or about other traders, market prices can lead to accurate forecasts of an asset's value. Plott (2000) formulated that prediction markets tend towards a "crystal ball equilibrium", a market that provides a forecast that is significantly better than one available from public sources of information. Prediction markets perform three tasks: they provide incentives for truthful revelation, they provide incentives for research and information discovery, and they provide an algorithm for aggregating opinions (Servan-

Schreiber et al., 2004).

Market participants are not all equal: In a landmark article, Black (1986) separates traders into informed traders and noise traders. Informed traders trade rationally, but noise traders nevertheless provide necessary market liquidity. Plott and Sunder (1982, 1988) show that markets have the ability to disseminate information from informed traders to uninformed (noise) traders. Additionally, markets aggregate information so that as a consequence, the market price reflects all information that is available to traders as required by the theory of rational expectations. In an efficient prediction market, the market price of the asset will be the best predictor of the event that is to be forecast (Wolfers and Zitzewitz, 2004). The authors also put forward the thesis of marginal trades - they state that it is not required for all individuals in the market need necessarily to be rational, so long as the marginal trade is performed by rational (informed) traders. The theory of marginal trades, however, is not accepted by all, most publicly Surowiecki (2004), who thinks that the marginal trader is a myth, as no individual or subgroup in a market can influence prices in the way the marginal-trader hypothesis suggests.

What is Special about Prediction Markets?

After all the evidence assembled by researchers for the efficiency of markets in aggregate, it should not seem strange that a market would be the best tool for forecasting events. In forecasting via expert opinion or individual polling, people will answer what they personally think or hope will happen in the future, or vote according to personal preferences or party affiliations. Markets have the advantage that people forecast or vote by discounting what they expect will really happen, their own hopes or affiliations retreating into the background.

Despite evidence of affiliation bias or home-team bias (Wolfers and Zitzewitz, 2004; Rhode and Strumpf, 2004), nor the fact that participants in a prediction market do not form representative samples, the market's invisible hand is deemed more dependent on a core of motivated marginal traders (Forsythe et al., 1999) and do not require large numbers of participants to function effectively. Participants are rewarded for accuracy in prediction, this is what enables unbiased outcome probabilities to be achieved (Rosenbloom and Notz, 2006). For example, it was shown that despite the participants belonging to a very narrowly defined socio-economic class, on the Iowa Electronic Markets the results of the market forecast in 1988 were not biased toward the presumed political leanings of the participants. (see Stix, 2008; Ho and Chen, 2007).

Ever since the University of Iowa (www.biz.uiowa.edu/iem) established the first prediction market with ongoing success - Berg et al. (2008) report that the Iowa Electronic Market has outperformed 964 traditional polls since 1988 74% of the time. Considerable evidence has been assembled to prove that prediction markets generate effective probability forecasts (see e.g. Berg et al., 2000, Berg and Rietz, 2003, Forsythe et al., 1999 or Surowiecki, 2004). Not only have prediction markets performed well in academic settings or election polling, but Sunstein (2006), Cowgill et al. (2009) and Healy et al. (2010) report that firms such as Google, Microsoft, Eli Lilly, Goldman Sachs, Deutsche Bank, Intel and Hewlett-Packard have already used prediction markets for in-house forecasting.

Ho and Chen (2007) show that prediction markets address major problems of survey-based and expert-based forecast approaches:

1. Participants are compensated for accuracy in forecasts
2. All participants can learn from others via the prediction market and can update their beliefs and develop a better forecast
3. The price discovery process weights accurate information more heavily and removes redundant and dependent information sources
4. Prediction markets can continue to integrate large numbers of participants at minimal incremental cost

The Challenges for Prediction Markets

A prediction market provides several other benefits when companies utilize them (Siegel, 2008): The market provides a constant gauge of risk probabilities that they would not otherwise track specifically or that often. The market provides a company with a level of transparency that would never have been achieved before; management is able to understand what employees really think, as they are providing their forecasts on an anonymous basis. Collaboration is achieved without social hindrances or many of the ever-present cognitive biases apparent in group processes. The awareness of the topic is greatly enhanced due to the unique status of the prediction market, individuals are induced to think differently about the outcome probabilities and trade accordingly. Lastly, in similar fashion to open-source projects such as Wikipedia or Linux or social media such as Facebook or YouTube, the lack of editors or gatekeepers provides users the chance to generate content on their own.

Whether the transferral of prediction markets from forecasting on large scales to narrowly defined corporate objectives has not been fully answered. Corporate prediction markets involve fewer traders, address far more complex problems, and management might require much more detailed data than provided by the typical prediction market. Healy et al. (2010) find that the double auction method is viable when the number of outcome states is small in comparison to the number of participants (thin markets: the number of contracts are too large compared with the number of traders) and when inverting beliefs into aggregated signals is not complex. In thin market cases, they find that other methods perform better. However, they look at extremely thin markets - they form markets with 3, 7 and 12 participants, respectively. They find that as soon as the number of participants increase, the standard errors of the forecasts decrease dramatically. Huber (2002) recommends only forming prediction markets with more than 25 participants in order to avoid the mentioned limitations.

According to the type of forecast that is required, different types of contracts are usual in prediction markets. In vote-share markets, participants try to forecast the share of the popular vote that a candidate will achieve. Winner-take-all markets depend on a binary outcome - contracts are specified to pay out if an outcome is achieved, e.g. if a candidate wins. If the outcome does not come to pass, the contract pays zero (see e.g. Gruca et al., 2005).

One issue that often arises is the question of whether for-real money markets perform better than play-money prediction markets. According to Hanson (1999), the common belief among economists is that markets should produce better forecasts if participants are obliged to risk real money. Real money is supposed to yield better incentives for information discovery. One drawback of this is that in a real market such as international stock markets, the weight of an individual's or institution's opinion is expressed by the amount of money they are willing to wager. On a prediction market, most participants will not have the kind of money a hedge fund has, and the weights that opinions receive are limited by the amount of (restricted) wealth that participants have. Strengths of opinions (in any direction) can be expressed on play-money markets just as effectively, as the strength of the judgments are not restricted by real wealth distributions. Also, setting up for-real-money prediction markets can be prohibitively difficult due to regulatory rules regarding betting and fiduciary issues, in addition to the significant costs and legal risks. Thus, research has found that play-money markets work very effectively (Servan-Schreiber et al., 2004; Rosenbloom and Notz, 2006); additionally, they find that psychological factors also play an important role (non-monetary incentives, attention derived from public rankings, etc.).

Prediction markets can also have limitations: Hanson (2007) finds that if the group of participants is too small or the incentives that were set for participants are too low to encourage active trading. If liquidity is not sufficient, the prediction market will not achieve the desired results. Speculative bubbles can form in experimental markets in the same way as on real markets (for an overview, see Sunder, 1995). Plott and Sunder (1988) find that speculative bubbles have also been observed in prediction markets. Also, in similar fashion to real markets, prediction markets are not completely immune to market manipulations. On the other hand, Rhode and Strumpf (2004) find evidence that prediction markets are surprisingly resistant to manipulations; prices returned quickly to their previous levels after initial changes due to manipulation.

According to Ray (2006), the conditions required for accurate forecasts on prediction markets are that a diversity of opinion must exist regarding future values of the underlying variables. A mechanism must exist whereby participants are able to express their opinions individually and without deferring to the majority opinion ("herding"). Prediction markets are set up so that each individual can provide their opinion by buying or selling the asset according to their judgment without regard to what others are doing; it provides the opportunity to actively take advantage of perceived inefficiencies, incorrect aggregate opinions or arbitrage possibilities.

If their judgment turns out to be accurate, they profit handsomely from trading against the opinions of the majority. Market participants must also be easily able to convert their private information to actions on the market - this is fulfilled by prediction markets, as participants are usually able to bet as much money as they desire, thus giving the strength of their convictions as much weight as the individual desires.

Setting up a Prediction Market

Siegel (2008) lists several key performance factors important for the success of a prediction market:

1. Find a vendor offering a program that allows you to quickly get a prediction market up and running
2. Ensure that the application is easy to use and understand
3. Ensure that the application also offers the necessary level of security to protect sensitive data

4. Identify specific questions for a marketplace
5. Make sure there is an internal administrator of the marketplace responsible for managing the software, providing support and communicating results
6. Recruit internal "champions"
7. Identify a broad range of traders to participate
8. Define which incentives participants will receive - not necessarily financial prizes, but often soft incentives can work the best

Even though Siegel provides his instructions on how to set up a prediction market for a corporate opinion poll, independently we incorporated all of his key factors into the setup of our marketplace.

A substantial number of participants in election stock markets exhibit less than rational behavior. Data from the 1988 Iowa Presidential Stock Market (Forsythe et al., 1992), the 1993 (Canadian) UBC Election Stock Market (Forsythe et al., 1995) and other election stock markets demonstrate that, on average, traders tend to exhibit some substantial judgment biases. In particular, traders' preferences over parties or candidates tend to color their perceptions, creating a "wishful thinking" effect (Forsythe et al., 1999). One of the aims of the current research is to analyze whether behavioral biases can be observed in individual participants, and if so, whether an experimental market can be "efficient" in aggregate, despite the individual biases and "inefficiencies". Pennock et al. (2001) find in an empirical study of the forecasting ability of prediction markets that prices strongly correlate with observed outcome frequencies.

2.2.3 Market Price Efficiency

If a price time series is efficient in the sense of the rational expectations hypothesis, then either the asset price or its natural logarithm should follow a random walk or a random walk with drift (Brooks, 2008). The majority of financial time series prices are non-stationary and contain a single unit root. By taking the first difference of the time series of prices, stationarity can be induced, implying that the log-differences (i.e. log-returns) follow a random walk. By this nature, the log-returns are thus unpredictable or only predictable to their long-term average value.

As defined by standard textbooks on time series analysis such as by Cowpertwait and Metcalfe (2009), a process is termed an autoregressive process AR(p), if the current observations y_t can be explained by past observations from the same time series. If a non-stationary time series y_t must be differenced d times before it becomes stationary, it is integrated of order d (for a description of necessary conditions for orders of integration, see e.g. Mills and Markellos, 2008). Applying the difference operator Δ once leads to an $I(0)$ process, i.e. a process with no unit roots that is stationary:

$$\text{If} \quad y_t \sim I(d), \quad \text{then } \Delta^d y_t \sim I(0) \tag{2.1}$$

The random walk, as the simplest case of an AR(1) process, is defined as:

$$y_t = \phi y_{t-1} + w_t, \quad \text{where } w_t \sim WN(0, \sigma^2) \tag{2.2}$$

where y_t are the current observations, ϕ is the (first-order) autoregression coefficient, and w_t are the residual white noise terms distributed with a mean of zero and a variance of σ^2 (see e.g. Carmona, 2004).

To test whether a time series is stationary, there are several methods that are applicable, falling into the categories of unit-root tests on the one hand or stationarity tests on the other hand (for an overview, see e.g. Cromwell et al., 1994, or Hamilton, 1994). The first method is to test for a unit root or non-stationarity of a time series - the null hypothesis of the test is that the series is non-stationary; this type of tests includes (augmented) Dickey-Fuller tests (Dickey and Fuller, 1979, 1981), the Phillips-Perron test (Phillips and Perron, 1988) or the Variance Ratio test (Lo and MacKinlay, 1988, 1989).

Probably the most popular method for unit-root testing is the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979, 1981). If the underlying financial time series of returns follows a random walk, an ADF test must fail to reject the null hypothesis of non-stationarity (a unit root exists). The same argument follows for the other unit root tests mentioned above. These tests work well for processes with clearly recognizable unit roots where the coefficient is not close to one.

Unfortunately, Cochrane (1991) states that research by authors such as Fama and French (1988), Huizinga (1987), Campbell and Mankiw (1987) and his own (Cochrane, 1988), have proved that many financial and macroeconomic time series (e.g. foreign exchange, stock prices, or GNP) *"...are all potentially of the 'borderline' type. First* *differences of these series have large positive autocorrelations for the first few lags, and then a series of small negative autocorrelations at very high lags...".* By "borderline type", Cochrane means that the process is stationary but with a root close to the non-stationary boundary. The probability of incorrectly failing to reject the null hypothesis of a unit root is thus quite high and renders the power of the test very low (see Brooks, 2008).

Due to the construction of the null hypothesis (the time series under observation is asssumed to be non-stationary) under the Dickey-Fuller tests, the null hypothesis of a unit root is never accepted, only not rejected. If a process is stationary with a root close to the unit circle, the power of these test can be low (see e.g. Cryer and Chan, 2008).

Stationarity tests, on the other hand, reverse the null and alternative hypotheses and test whether the process is stationary under the null hypothesis. One such test is the KPSS test (Kwiatkowski et al., 1992). Therefore, if the time series is constructed as observed by Cochrane (1988), Fama and French (1988), Huizinga (1987), or Campbell and Mankiw (1987), testing under the null hypothesis of stationarity rather than the opposite is of distinct advantage. For this reason we follow the strategy of testing for stationarity via the KPSS test.

As mentioned above, the first step is to perform the test for stationarity. If the prices (or log-prices) of time series prove to be non-stationary, then take the first difference and test again.

The next step is to test the (stationary) differenced series for autocorrelation. In order for the process to conform to a random walk process, the residuals must be white noise (see Equation 2.1). The standard test for autocorrelation can either consist of graphically viewing the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) correlogram graphs, and/or performing a formalized test for autocorrelation of the residuals (see e.g. Brooks, 2008). Under the graphial alternative, the ACF correlogram shows the autocorrelations for each lag k. The PACF correlogram shows the correlation that results after removing the effect of any correlations due to the terms at shorter lags. For example, the PACF of an AR(1) process will be zero for all lags greater than 1. The correlogram for a random walk shows a gradual decay from a high serial correlation as a common feature (see e.g. Cowpertwait and Metcalfe, 2009).

In order to determine statistical significance of (non-)stationarity, a formalized test must (also) be performed in adddition to visual observation of the presence of non-stationarity in correlograms. One such standard test is the Durbin-Watson test for autocorrelation (Durbin and Watson, 1951). Under the null hypothesis, the errors at time t - 1 and t are independent of one another. If the null hypothesis cannot be rejected for the time series, then we have ample evidence of stationarity of the first differences and thus of the original price series conforming to a random walk. This, in turn, confirms weak-form price efficiency.

2.3 Research Objectives

In order to test for outcomes and interactions of participants in a financial market, it is useful to isolate as many parameters as possible in order to study the desired effects without noise clouding potential result interpretation. As real-life financial markets are very complex structures with many participants in differing legal and environmental settings, an experimental market can function as a proxy for such a market. Prediction markets have been shown to provide excellent results as experimental markets that model real-world market activity (see e.g. Wolfers and Zitzewitz, 2004; or Berg and Rietz, 2003). As such, a prediction market provides a unique platform for testing human decision making in a controlled financial market setting.

If the experimental market fulfils the basic economic roles of a financial market (interactions of buyers and sellers determine the price of an asset via the price discovery process, financial markets provides liquidity, search and information costs are reduced; see e.g. Fabozzi and Modigliani, 2003) it can be considered a good proxy for a financial market. The economic roles of a financial market are fulfilled if prices reflect the aggregate information collected by all market participants, i.e. the market is efficient.

As a form of experimental markets, an on-line electronic trading platform was established as a prediction market which allowed study participants to trade on the outcomes of soccer games during the European Soccer Championship 2008. The platform admitted students at two universities in Vienna (34% of the participants), as well as employees (29%) and clients (16%) of an international consulting firm and other professionals (21%).

The broad professional and demographic distribution provided a realistic environment for trading stocks as one would in real-life markets. Supply and demand determine prices, investors have to deliberate the consequences of posting bid or offer trading indications and the price discovery mechanism condenses the varied information into one figure, the current price.

In order to test whether the prediction market used in the current research fulfils its role as a proxy for other financial markets, we test it for forecast effectivity, in order to determine whether available information was aggregated into the quoted prices, and for price efficiency in the weak-form sense (Fama, 1965).

On the current prediction market, of interest is:

1. Whether the soccer prediction market offers reliable predictive values for the final winner and for the development of teams that did not win,
2. Whether participants in the market show behavioral biases on an individual level,
3. Whether the market is efficient in aggregate in a weak-form manner or not,
4. Whether the market can be used as a proxy for real-life markets, and
5. Whether the results of the behavioral biases can be set in a risk management context in order to recognize such biases before they become problematic.

As a consequence, we formulate and test the following hypotheses:

Hypothesis 1A: *The soccer prediction market offers good predictive values for the final winner of the competition.*

Hypothesis 1B: *The soccer prediction market offers good predictive values during the competition for non-winning teams.*

Hypothesis 2: *Participants in the prediction market show evidence of behavioral biases* (see chapter 3).

Hypothesis 3: *The soccer prediction market is efficient in aggregate despite potential behavioral biases of individual participants.*

Hypothesis 4: *The soccer prediction market can be used as a proxy for real-life markets.*

2.4 Method

2.4.1 Participants

The raw sample of participants comprised 434 participants who provided the information required for study participation (personal details, login details, and a brief questionnaire). Of all the participants in the study sample, 340 actively traded while 94 participants did not perform any trades, thus the 340 active participants formed the final study sample. Of this sample, most participants were male (62%) and between the ages of 20 and 29 (56%). Professionally, the sample was split into one-third students, one-third employees of a major consulting company (the company sponsored the prediction market) and the other third consisted of clients of the consulting company, as well as friends and acquaintances.

2.4.2 Research Design

Initial Setup of the Trading Platform

A prediction market was set up as an online trading platform on the European Soccer Championships 2008. Each of the 16 teams that contested the finals of the Championships was allotted a stock, and each participant who signed up for the market received an initial basket of all 16 stocks (the same number from each team) as well as online trading cash. Trades were performed online as a continuous double auction (trades were constantly updated and orders were matched if bid and offer prices with matching prices were entered). The current price at each moment in time constituted itself from the current matched trades. The order book was open, meaning that all participants could see all outstanding unmatched bids and offers (prices and quantities).

For 2 Euros each, participants signed up to trade and were assigned 10,000 units of cash (football cents) and additionally a stock basket containing 100 units of each team's stock (e.g. Austria, Germany, Netherlands, etc.) also worth 10,000 cash units (the total of 20,000 cash units were equivalent to 2 Euros). Participants who signed up prior to the actual begin of the European Championships received 16 teams in a stock basket. Each stock started out at a price of 10,000 currency units / (16 teams * 100 shares) = 6.25 currency units per share, the price then fluctuating, determined only by supply and demand.

The market was structured as a "winner-takes-all" market, meaning that the one team that ultimately won the championship would redeem at 100% (10,000 currency units), and all other teams at 0. This potential price fluctuation between 0 and 100 meant that the price at any given point in time could be interpreted as a probability of that particular team winning the trophy. Exactly at inception, each team was thus assigned a 6.25% chance of winning, which quickly changed following the initial trades (arbitrage was possible for a very short period at that time before prices settled down to an initial equilibrium).

This procedure has been effectively used by the Iowa Electronic Markets, one of the pioneers of on-line prediction markets (see e.g. Forsythe et al. (1992), or www.biz.uiowa.edu/iem). To make things easier, the consulting company sponsored the 2 Euros per participant, and the results were donated to charity after the event.

Trading Details

Those participants who signed up later during the course of the tournament, or purchased further stock baskets after individual teams had been knocked out of the tournament, received stock baskets comprising fewer than the initial 16 teams (the price of a stock basket remained constant at 10,000 currency units). The prices for each of the remaining teams adjusted accordingly. If, for example, 6 of the 16 teams had already been eliminated, each team's share price in a stock basket comprising the 10 surviving teams was thus 10 rather than the initial price of 6.25 when all teams were in the running. This corresponds to the fact that these surviving teams obviously also had a higher chance of advancing further towards the finals!

The on-line trading platform ensured that investors could trade at any moment in time 24 hours a day. The number of trades able to be placed was theoretically unlimited. Participants were able to buy additional cash units or stock baskets at any point in time on-line. In order to avoid potential regulatory issues (futures markets fall under supervisory requirements by national regulatory authorities and require a license, and public betting lies within the exclusive domain of the state gambling monopoly), participants were informed before in the general terms and conditions of the platform that the two Euros initial investment would be donated by a sponsoring company and would be donated to charity after the event, and that any losses incurred over and above this sum resulting from additional cash purchases would be invoiced after maturity of the market with the request that the sum be donated to the specified charity organization.

Investors were able to see their total portfolio at any point in time and could make trades directly from the on-line interface to their portfolio (see Figure 2.2).

Figure 2.2: Soccer Prediction Market Portfolio Excerpt

	ID	Art	Anzahl	Preis	Kosten	Datum ▼
H	ESP	v	100	57.99	5799.00	29.06.2008 21:10
H	GER	v	300	48.01	14403.00	29.06.2008 20:36
H	ITA	r	47	0.00	0.00	22.06.2008 23:23
H	ITA	k	47	17.65	829.55	22.06.2008 21:53
H	NED	r	200	0.00	0.00	21.06.2008 23:17
H	CRO	r	200	0.00	0.00	20.06.2008 23:36
H	GER	k	200	22.00	4400.00	20.06.2008 18:17
H	NED	k	100	42.89	4289.00	20.06.2008 18:17
H	TUR	v	100	1.50	150.00	20.06.2008 18:16
H	CRO	k	100	12.96	1296.00	19.06.2008 22:53

Note: The figure above shows an excerpt from a portfolio containing: in column ID: the country/team shortname in 3-digit ISO format, column Art: the type of transaction, v for a sale order, k for a purchase order, r for a repurchase at maturity; column Anzahl: the number of stocks in the transaction; column Preis: the transaction stock price; Kosten: the total cost of the transaction calculated as price x stock number; column Datum: the transaction date and time.
The teams' names are the following ISO codes: AUT: Austria, CRO: Croatia, CZE: Czech Republic, ESP: Spain, FRA: France, GER: Germany, ITA: Italy, NED: the Netherlands, POL: Poland, POR: Portugal, ROU: Romania, RUS: Russia, SUI: Switzerland, SWE: Sweden, TUR: Turkey.

The order book for each team was open, so each participant could see all bids and offers available for each team and trade correspondingly. Returns were calculated daily and posted on-line (see Figure 2.3) in a ranking table for all to see (trading returns were calculated as realized returns on a simple non-annualized percentage basis as there was only one trading period, i.e. a period return was sufficient for comparison purposes. Revaluations of portfolios with unrealized returns were not calculated).

After the Final

In addition to the honor of having won, the most successful traders (those who achieved the highest cumulative return) were presented with prizes after the tournament had ended[3]. Both of these facts acted as powerful motivators for participants to trade, as evidenced by over 50,000 trades posted within the 3-week period; an additional motivator was the Activity Index indicating the cumulative number of matched trades for each investor along with his or her current ranking by activity.

[3] The first in the category "Student" won a semester's tuition for free, the winner of the category "Consultant" won a wellness weekend, and the winner of the category "Client" won a trip to the country of the winning team.

Figure 2.3: Soccer Prediction Market Ranking Snapshot

Ranking vom 06.07.2008 12:00

Pos.	User	Return ▲	Volumen	Aktivitaet	Spenden
125.	no_nickname	+3.00% (125.)•	599.00 (347.)•	2 (341.)•	0.00 (74.)•
126.	Börsenaufsicht	+2.61% (126.)•	575.00 (348.)•	1 (351.)•	2000.00 (72.)•
127.	steffi	+2.37% (127.)•	24761.05 (210.)•	36 (213.)•	10000.00 (62.)•
128.	Kerstin	+2.17% (128.)•	1007871.29 (19.)•	266 (51.)•	980000.00 (1.)•
129.	jacqui	+1.86% (129.)•	2097.00 (338.)•	8 (318.)•	0.00 (74.)•
130.	mreichelt	+1.80% (130.)•	10359.40 (285.)•	27 (250.)•	0.00 (74.)•
131.	Clemi83	+1.10% (131.)•	220.00 (351.)•	1 (351.)•	0.00 (74.)•
132.	7kiwi	+1.03% (132.)•	31834.55 (170.)•	15 (291.)•	0.00 (74.)•
133.	TTMM	+0.57% (133.)•	115.00 (354.)•	1 (351.)•	0.00 (74.)•
134.	Greece2008	+0.44% (134.)•	22620.81 (217.)•	39 (203.)•	0.00 (74.)•
135.	mhofner	+0.38% (135.)•	13575.45 (262.)•	19 (281.)•	0.00 (74.)•
136.	"The Agent"	+0.26% (136.)•	52.00 (357.)•	1 (351.)•	0.00 (74.)•
137.	teamchefin11	+0.21% (137.)•	42.00 (359.)•	2 (341.)•	0.00 (74.)•
138.	yildi	+0.21% (138.)•	81.50 (355.)•	2 (341.)•	0.00 (74.)•
139.	Basler	+0.20% (139.)•	10540.00 (284.)•	12 (301.)•	0.00 (74.)•

125-139 von 361 < Zurück | Vor >

Note: The figure above shows a snapshot of a portion of the online published ranking of the participants showing: in column Pos: the absolute ranking number of the participant; column User: the user name of the participants; column Return: the total return to date per participant; column Volumen: the total trade volume per participant to date and the ranking by volumen in parentheses; column Aktivitaet: the activity per participant measured as the number of trades to date and the ranking by activity in parentheses; column Spenden: the sum and ranking in parentheses of each participant's additional cash purchases, indicated as a post-event donation to charity.

Contract and Price Description

Participants traded stock on the 16 finalist teams of the European soccer championships and transacted over 50.000 trades over a 3-week period. The market was structured in the form of a futures market with winner-takes-all contracts. Participants received initial stock and trading cash in equal amounts with which they could immediately start trading. Prices were determined by online trades with an open order book as a continuous double auction. Participants were always able to perform the following actions:

- Buy further baskets of all available stock (the number of different stocks in the basket depended on the time of purchase - the basket only contained those stocks whose teams were still in the tournament. If 1 unit of stock basket worth 10,000 monetary units was purchased (equivalent to 1 Euro), this was divided among the number of outstanding teams. An example: While all 16 teams were in the tournament, a basket consisted of 10,000 / 16 = 625 stock per team. After the round-robin pool stage, only 8 teams were left; the basket was updated to include 10,000 / 8 = 1,250 stocks per team).

- Buy or sell individual stocks as market or limit orders

- Cancel existing unfilled orders
- Increase cash holdings

The initial combined price for a stock basket was 100 units. All losing teams‘ prices converged to a payout of 0, and the one winning team‘s price converged to a payout of 100.

2.4.3 Data and Measures

Transaction Data

The original data consisted of a file containing 50,047 individual lines of transactions. These contained the following order categories:

- Matched orders (including buyer ID, seller ID, underlying team, price, quantity, time), type “n”
- Cancelled orders, type “cv” or “ck”
- Initial stock basket transactions (at inception for each participant), type “i”
- Initial cash transactions (at inception for each participant), type “c”
- Additional stock basket purchases, type “k”
- Additional cash purchases, type “v”
- Final stock repurchases (at either zero or 100 at maturity of the contracts), type “r”

From these transactions, incorrect or partially missing data were cleansed, cancelled orders were removed and all transactions from one participant were removed who had proved to have cheated and violated the general terms and conditions of the market. From each order containing a buyer and a seller, the price of the transaction was extracted and used to form each of the 16 teams’ price series over time; This resulted in a file containing 48,757 individual transactions. Additionally, the price and order quantity was added to each participant’s portfolio at each point in time, depending on whether the participant has bought or sold the particular quantity.

Standardization of Price Series

As neither the length of time each team was in the competition was equal (teams dropped out at both round-robin and knockout stages, or it became clear that the team was not able to progress at different times), nor the number of trades constituting a price series were equal, standardization was necessary. We standardize the price series to convert from such individualized high-frequency series to standardized time intervals. Equal numbers of data points at identical points in time are necessary to be able to perform statistical tests and perform comparisons between price series, as this is non-trivial for high-frequency data of varying lengths and densities of price movements.

As a standardization rule, time frames had to be selected in which neither too many nor too few price changes took place (the former would lose price information under standardization, the latter would form ladder functions). Under these criteria, we selected a decimal time interval of 100 observation points per day, resulting in 2,702 individual data points, beginning at a decimal date of 3.49 (in decimal form, slightly before midday on the third of June, when the first games began) to decimal date 30.50 (exactly midday on June 30), after the last redemption trade was executed (the decimal point 0.50 is midday if there are 100 observation points beginning at midnight and continuing for 24 hours).

Prediction Quality

Obviously, on a prediction market, one of the points of interest is if, and how well, the market managed to predict winners and losers, and how far in advance this was achieved. In order to determine prediction quality within a confidence level of 99%, we find the length of time a team's price was below a price of 1 (indicating the market assigns a 1% chance of winning), counting backwards from the end of the team's price series. We then put this in relation to the entire length of its price series.

If a team that ultimately lost (i.e. 15 of the teams) has a price of below 1 for a long time before it is finally eliminated from the tournament, that will indicate much better prediction quality than if the team's price was very high for a long time and only suddenly crashed to zero at the end. Obviously, the converse is true for the team that ultimately won. The methodology is outlined in the following steps.

We analyze the prediction quality of the teams as individual teams and in aggregate over more than one team with the following approaches:

1. Each team's price can range in the interval between 0 and 100. As outlined above, the time duration of the prediction market was divided into 2,702 equal decimal time increments; the beginning of the market was at decimal date 3.49 and the end at decimal date 30.5, 2,701 decimal time increments later. The exact time increment is located at which the team's price is equal to or below a price of 1 for the 10 following time increments (or until the contract of the team matures and the contract is redeemed at price zero)[4].

   An example: A team drops below a price of 1 for 10 consecutive time points at decimal date 10.27, this corresponds to 679 time increments after the beginning date of 3.49. The contract is redeemed at price zero, i.e. the team drops out, at decimal date 16.99, or 1,351 decimal time increments after the beginning. This means that the team's price was below 1 for 1,351 - 679 = 672 decimal time steps.

   Then the decimal point in time at which the price drops below 1 is taken as a percentage of the total number of time increments for the team.

   Continuing above example: the 672 time steps in which the team's price was above 1 are related to the entire time (1,351 time steps) that the team was in the tournament: 679 / 1,351 = 50.26% of the time.

   Lastly, this percentage is subtracted from 1 to give a number which shows for how long the prediction was within a 99% confidence interval over all possible prices.

   Continuing above example: the 672 time steps where the team's price was below 1 are related to the entire time (1,351 time steps) that the team was in the tournament: 672 / 1,351 = 49.74% of the time.

   This indicates that the market was of the opinion for almost one half of the time the team was in the tournament that there was less than a 1% chance of the team winning the tournament, an excellent indication of prediction quality as measured within a 99% confidence interval.

2. The development of the teams' prices are visualized in a graph

3. The mean squared error (MSE) of the teams' prices are contrasted with the price development

[4]This methodology is performed for all non-winning teams, the methodology for the winning team, whose price converges to 100 rather than zero, is applied analogously, the number of timesteps are counted that the team is above 99 rather than below 1.

Market Liquidity

As outlined in the Introduction section, liquidity issues are of vital importance to a prediction market. On the one hand, if there are not enough participants and/or too many contract types, then thin markets are to be expected (see Healy et al., 2010). As we have 340 actively trading participants, 16 contracts and over 50,000 trades, thin markets are not to be expected. However, price liquidity, that could result in too broad bid/offer spreads as outlined above, could have a negative impact on prediction quality.

In order to research market liquidity, we draw on the methodology initiated by Bao et al. (2011). They state that the simplest measure of illiquidity, the bid-offer spread, does not fully capture all aspects of illiquidity that occur. In particular, analyzing only the bid-offer spread misses market depth and resilience of the market, as well as transitory price movements (Bao et al., 2011). In order to capture the magnitude of these transitory price movements, they construct a measure of illiquidity γ for individual corporate bonds based on negative serial correlations. This approach goes back to Niederhoffer and Osborne (1966), who were among the first to document the relationship between negative serial correlation and illiquidity issues. Bao et al. (2011) define illiquidity as:

$$\gamma = -Cov(\Delta p_t, \Delta p_{t+1}), \quad \text{where} \tag{2.3}$$

$$\Delta p_t = p_t - p_{t-1} \tag{2.4}$$

is the price change of the observed price p_t from t - 1 to t with the assumption that the fundamental asset follows a random walk.

We test for illiquidity of each price series using this methodology.

Market Efficiency

We performed KPSS tests for each team's standardized price level series, for the first differences of each team's prices, and tested visually as well as formally with a Durbin-Watson test for each team's first differences' potential autocorrelation.

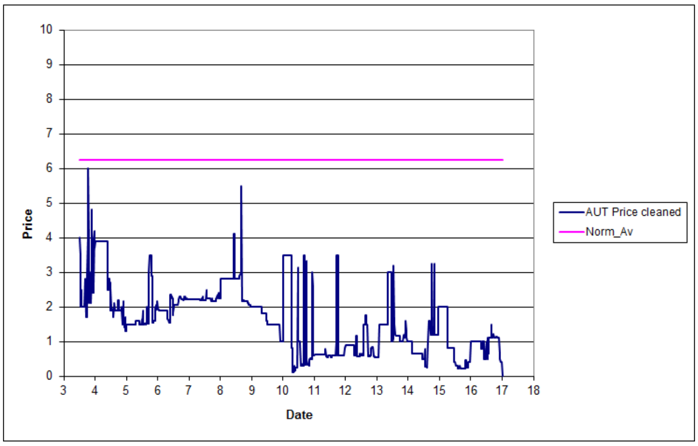
If all of the 16 price paths are random walks, we can conclude that the prediction market as a whole conforms in aggregate to weak-form price efficiency. In our case, we tested for weak-form efficiency, but not for semi-strong form efficiency, as the influence of news on price generation was beyond the scope of the experiment.

2.5 Results and Discussion

2.5.1 Prediction Quality

The first analysis of the prediction quality begins with a visual observation of the 16 price graphs. Figure 2.4 shows the development of the Austrian team, who did not make it past the pool stage of the tournament. Apart from initial excess volatility in the price finding phase, the development is quite stable. At the beginning of the tournament each team began at a price of 6.25 (as there were 16 teams, and the market was structured as a winner-takes-all futures market, 100 units were divided by 16 teams). This meant that initially there were arbitrage possibilities on the one hand when trading commenced, but these were quickly utilized and prices stabilized within the first day, which can be seen in Figure 2.4.

Figure 2.4: Price Development for AUT

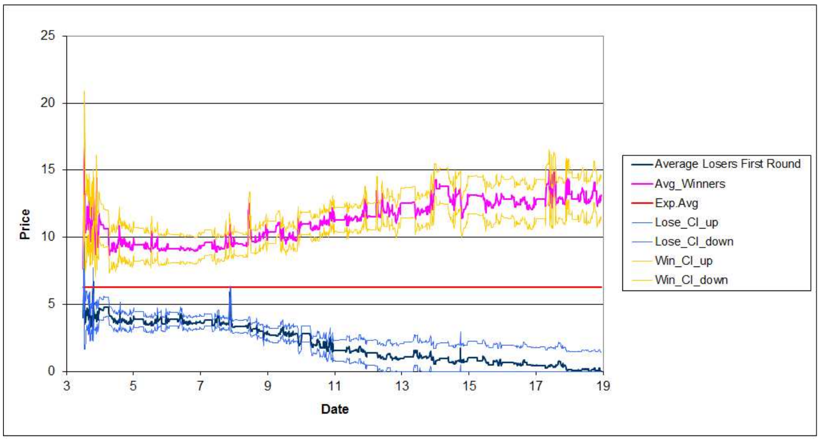


The prices of the teams quickly made their way to levels that reflected participants' judgments of their potential. As one can see in Figure 2.4, even though Austria was one of the two hosts (the other was Switzerland) of the tournament, there was no home-team bias to be observed, as the price never reached the objective level of 6.25 after trading started (the pink line in Figure 2.4).

The First Tournament Round

The first test of prediction quality is shown in Figure 2.5. The graph shows the teams in two groupings - the higher grouping (the pink line) represents the aggregated prices of those teams who later made it past the initial pool stage of the tournament. The lower line shows the aggregated prices of those teams who did not make it past the pool stage. The yellow and light blue lines show one standard deviation bands around the aggregated prices.

Figure 2.5: First Round Prices



Note: The graph shows the teams in the yellow/pink colors that made it past the first round to the knockout stages with a one standard deviation band up and down around the prices. The blue lines show the price development of those teams that lost in the pool stage, also with a one standard deviation band around the price series, bounded on the lower side by zero.

What is interesting is that even though the teams were not all forecast correctly during the beginning stages of the tournament as concerns their ultimate achievements, in aggregate the predictions showed great quality. The aggregated prices of the teams that won in the pool stage were always above those who did not, also including a standard deviation of error margin. The collective market opinion was settled quickly and represented the aggregated outcomes very effectively. In particular, the losing teams' aggregated price level dropped to a level of around or under 1 from day 12 onwards (with 7 days remaining in the pool stage).

If one takes the contract price as the market's percentage evaluation of an ultimate victory (which would pay out 100 units), then the level of 1 unit would represent a 99% confidence interval around zero that the team would ultimately lose. Thus we take this as proof that in aggregate, the market identified the losing teams extremely effectively, as they sank below this level just after half-way through the initial pool round of the tournament (at which stage no results were actually fixed yet, any team still had a theoretical chance of advancing). This we interpret as a proof of Hypothesis 1B.

Quality Check of all Teams

The prediction quality for the teams is checked with three different methods. First, we tabulate the 16 teams in Table 2.1 and calculate the percentage of time increments at which 1 (or 100) is breached at least for 10 subsequent time intervals.

Table 2.1: Results of Prediction Quality Analysis

Team	% Time Price > 1	1 - (% Time Price > 1)
AUT	50.26%	49.74%
CZE	83.60%	16.40%
FRA	75.43%	24.57%
GRE	56.08%	43.92%
POL	44.04%	55.96%
ROU	50.31%	49.69%
SUI	39.52%	60.48%
SWE	57.48%	42.52%
CRO	99.94%	0.06%
ESP†	99.96%	0.04%
GER	99.96%	0.04%
ITA	99.95%	0.05%
NED	99.95%	0.05%
POR	99.88%	0.12%
RUS	99.96%	0.04%
TUR	99.82%	0.18%
Mean	78.51%	21.49%

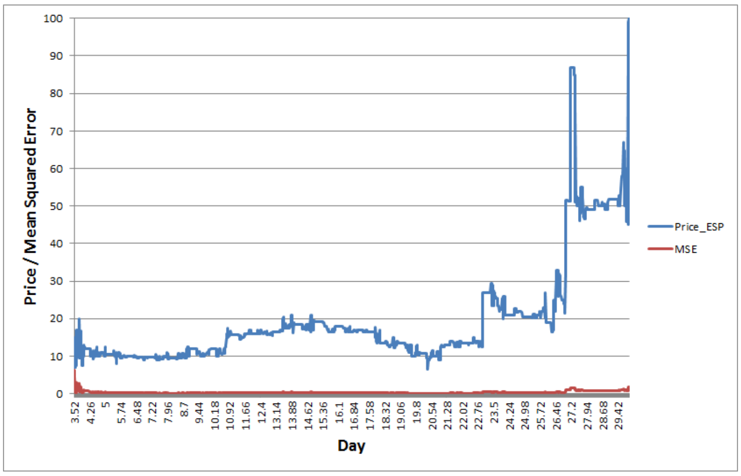
Note: †: The number of time steps for Spain is quoted as < 99 rather than > 1

What is interesting about the figures in Table 2.1, is that the prediction quality is good to excellent for those teams that did not survive the first round. The market quoted a price of 1 or less for continuous time periods ranging from 16.4% to 60.5% of the total time a team was in the running. The Romanian team was the team that was the best predicted of all.

The teams that progressed obviously were in much closer contention, and voting was close until the end, as can be seen in Table 2.1. In aggregate the market found close to the correct valuation for the last fifth of the contract periods, which would be considered good prediction quality, if it were not for the close calls among the teams that progressed.

In Figure 2.6 an example is shown for the development of the Mean Squared Error for the team ESP (that ultimately won the competition). It can be seen that the MSE decreases exponentially from the beginning of trading, which we interpret as a sign of good prediction quality regarding the individual teams and a further partial confirmation of Hypothesis 1B.

Figure 2.6: Development of MSE of ESP

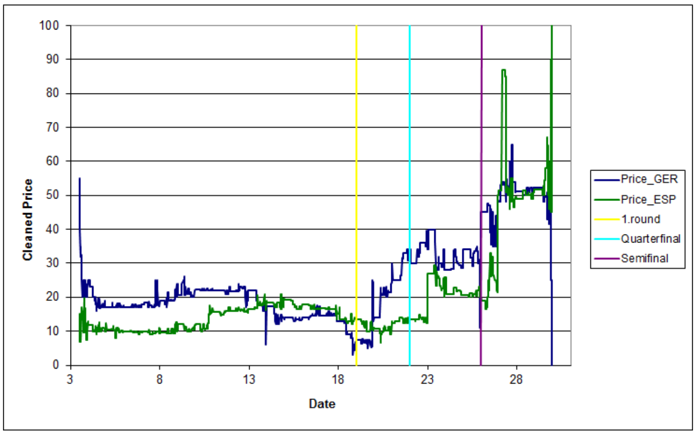


Note: The graph shows the price in blue and the Mean Squared Error in red on the same axis.

The Finalists

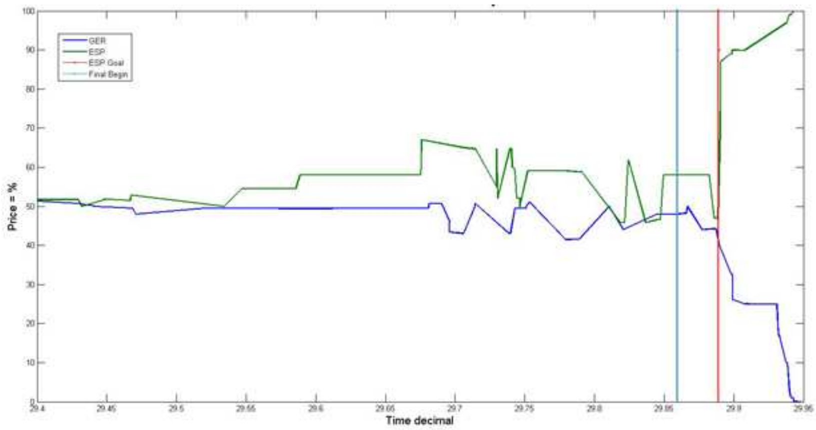
A visualization of the development of the two finalists shows what was already indicated in Table 2.1 - that it was a close call for all of the tournament.

Figure 2.7: The Winners: GER vs ESP



Germany (the blue line in Figure 2.7) begins the tournament as favored over Spain (the green line) until it loses a first-round match. It can be seen that publicly available information is rapidly incorporated into prices, even though a formal proof of semi-strong form efficiency is beyond the scope of this paper. The next crossover point is when Germany wins its quarterfinal match. From then on it is favored over Spain (the ultimate winner) by and large until the final day.

Figure 2.8: The Final Day: GER vs ESP



The final day is close the whole time until Spain shoots the only goal in the match. From then on the market rapidly adjusts to the collective judgment that Spain is going to be the ultimate winner. Despite the closeness of the contract prices for the two teams, if one had to choose between the two, the market was betting on Spain for almost the entire day in absolute value, as can be seen in Figure 2.8. We interpret this as final proof of confirmation of Hypothesis 1A, that the market incorporates information extremely quickly and provides good predictive quality throughout the tournament for the final teams. Figure 2.12 shows the development of the prices for 4 selected teams in the tournament.

2.5.2 Market Efficiency

In order to test Hypothesis 3, the test for price efficiency, we first test all 16 price series for stationarity. We test for weak-form efficiency as indicated by prices or returns conforming to a random walk. We perform the following steps:

1. Test each price series for log-price stationarity with the KPSS test
2. If a price series is not stationary, then calculate first differences
3. perform KPSS test for stationarity on log-differences
4. graphically analyze the sample ACF and PACF correlograms

5. perform the Durbin-Watson Test for autocorrelation

The results of these steps are shown in Table 2.2. The 16 teams' ISO country codes in column 1 are as follows:

- AUT: Austria,
- CRO: Croatia,
- CZE: Czech Republic,
- ESP: Spain,
- FRA: France,
- GER: Germany,
- ITA: Italy,
- NED: the Netherlands,
- POL: Poland,
- POR: Portugal,
- ROU: Romania,
- RUS: Russia,
- SUI: Switzerland,
- SWE: Sweden,
- TUR: Turkey.

Table 2.2: Results of Stationarity Tests

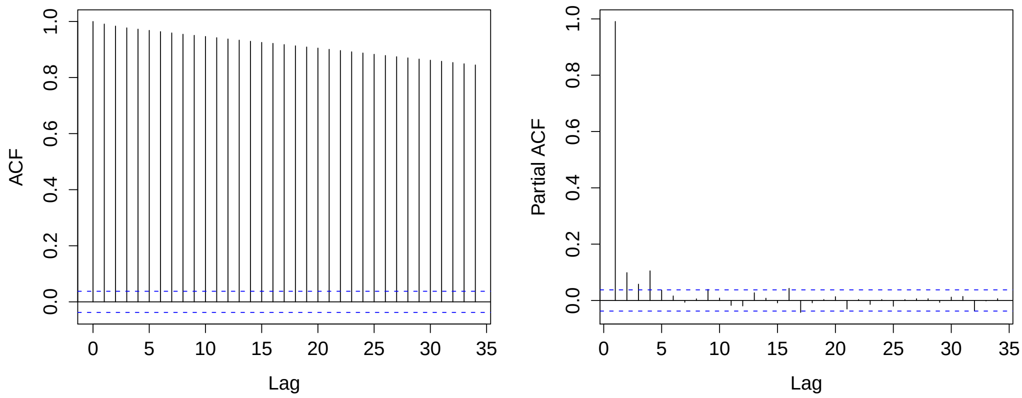
Team	logPrice / Diff.	KPSS Level	p-Value	Durbin-Watson Stat.	p-Value	Accept / Reject
AUT	logPrice	7.04***	0.01			Reject
CRO	logPrice	6.97***	0.01			Reject
CZE	logPrice	11.92***	0.01			Reject
ESP	logPrice	10.64***	0.01			Reject
FRA	logPrice	19.20***	0.01			Reject
GER	logPrice	11.51***	0.01			Reject
GRE	logPrice	12.37***	0.01			Reject
ITA	logPrice	4.02***	0.01			Reject
NED	logPrice	16.80***	0.01			Reject
POL	logPrice	9.49***	0.01			Reject
POR	logPrice	11.95***	0.01			Reject
ROU	logPrice	8.57***	0.01			Reject
RUS	logPrice	10.19***	0.01			Reject
SUI	logPrice	12.20***	0.01			Reject
SWE	logPrice	6.79***	0.01			Reject
TUR	logPrice	9.02***	0.01			Reject
AUT	Difference	0.07	0.10	1.93	0.23	Accept
CRO	Difference	0.21	0.10	2.27	0.76	Accept
CZE	Difference	0.16	0.10	2.04	0.23	Accept
ESP	Difference	0.10	0.10	1.90	0.76	Accept
FRA	Difference	0.08	0.10	2.01	0.99	Accept
GER	Difference	0.27	0.10	2.01	0.23	Accept
GRE	Difference	0.04	0.10	2.01	0.76	Accept
ITA	Difference	0.18	0.10	1.67	0.76	Accept
NED	Difference	0.39	0.10	2.04	0.77	Accept
POL	Difference	0.03	0.10	2.06	0.23	Accept
POR	Difference	0.38	0.10	1.99	0.23	Accept
ROU	Difference	0.12	0.10	2.05	0.76	Accept
RUS	Difference	0.17	0.10	2.04	0.23	Accept
SUI	Difference	0.02	0.10	2.07	0.23	Accept
SWE	Difference	0.21	0.10	2.03	0.76	Accept
TUR	Difference	0.08	0.10	2.03	0.23	Accept

Note: The top half of the table shows KPSS tests for stationarity of log prices for each team. The p-Values of the test are all significant, therefore the null hypothesis of stationarity must be rejected. The bottom half of the table shows KPSS tests for stationarity of the first differences of the log prices, all of which indicate that the null hypothesis of stationarity cannot be rejected. The Durbin-Watson p-Values show that the null hypothesis of no autocorrelation cannot be rejected, i.e. the first differences of the log prices are not autocorrelated.

*** denotes significance at the 1% level

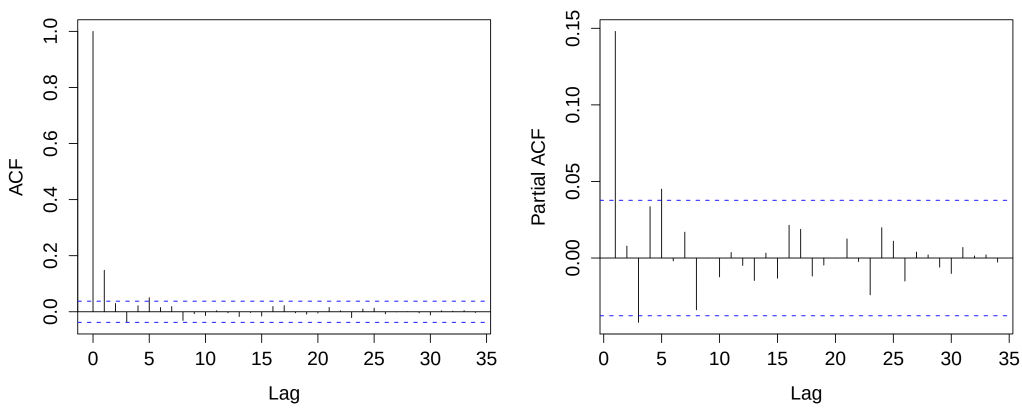
The p-values in Table 2.2 for the KPSS tests look similar for each block, but these are the true values that result from usage of the function kpss.test from the tseries package in R (R Development Core Team, 2011); the accompanying documentation states that p-values are interpolated from Table 1 of Kwiatkowski et al. (1992).

Figure 2.9: ACF and PACF of log Price Series for GER



Visual inspection of the ACF correlograms of the price series show a gradual decay from a high serial correlation - this is a notable feature of a random walk series (for one example see Figure 2.9). The PACF shows two marginally statistically significant values; they can be ignored, however, because they are small in magnitude. Additionally, at the indicated confidence level of 95% (the dashed lines in the graphs), about 5% of the values are expected to lie outside of the lines even when the underlying values are zero. Therefore, there is good evidence from the correlograms that the underlying price series follows a random walk (compare Cowpertwait and Metcalfe, 2009, pp. 36 and 75).

Figure 2.10: ACF and PACF of log-Difference Series for GER



The ACF and PACF graphs of the log-differences of the price series in Figure 2.10 confirm the assumption of a random walk in the levels (the original undifferenced price series) and show a different picture; the marginally significant autocorrelations in the PACF graph do not counter the findings of a random walk for the price series under the conditions outlined above.

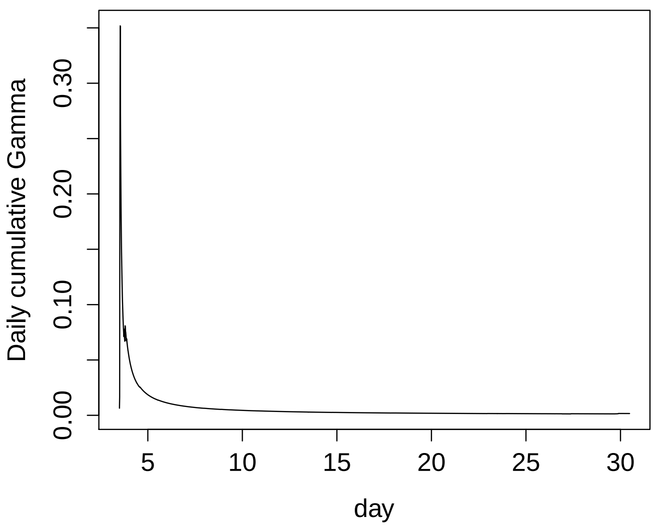
The results of the steps 1 to 3 and 5 are shown in Table 2.2. The KPSS test for stationarity rejects the null hypothesis of stationarity for all 16 log price series. The KPSS test on the differences of the log prices, however, shows that the price differences are stationary. The confirmation of a random walk process for all price series is given by the results of the Durbin-Watson test for autocorrelation: all 16 difference series confirm the null hypothesis that there is no autocorrelation to be seen.

Thus we can confirm Hypothesis 3 - the prediction market prices are random walks and confirm price efficiency in aggregate for the market.

(Il)liquidity Analysis

The liquidity analysis was performed using the Bao et al. (2011) gamma measure as explained above. The results are shown in Figure 2.11 for one of the teams as an example. All teams' log-prices were analyzed according to this methodology and show essentially the same behavior as shown in the example in Figure 2.11.

Figure 2.11: Gamma Function for ESP log Prices

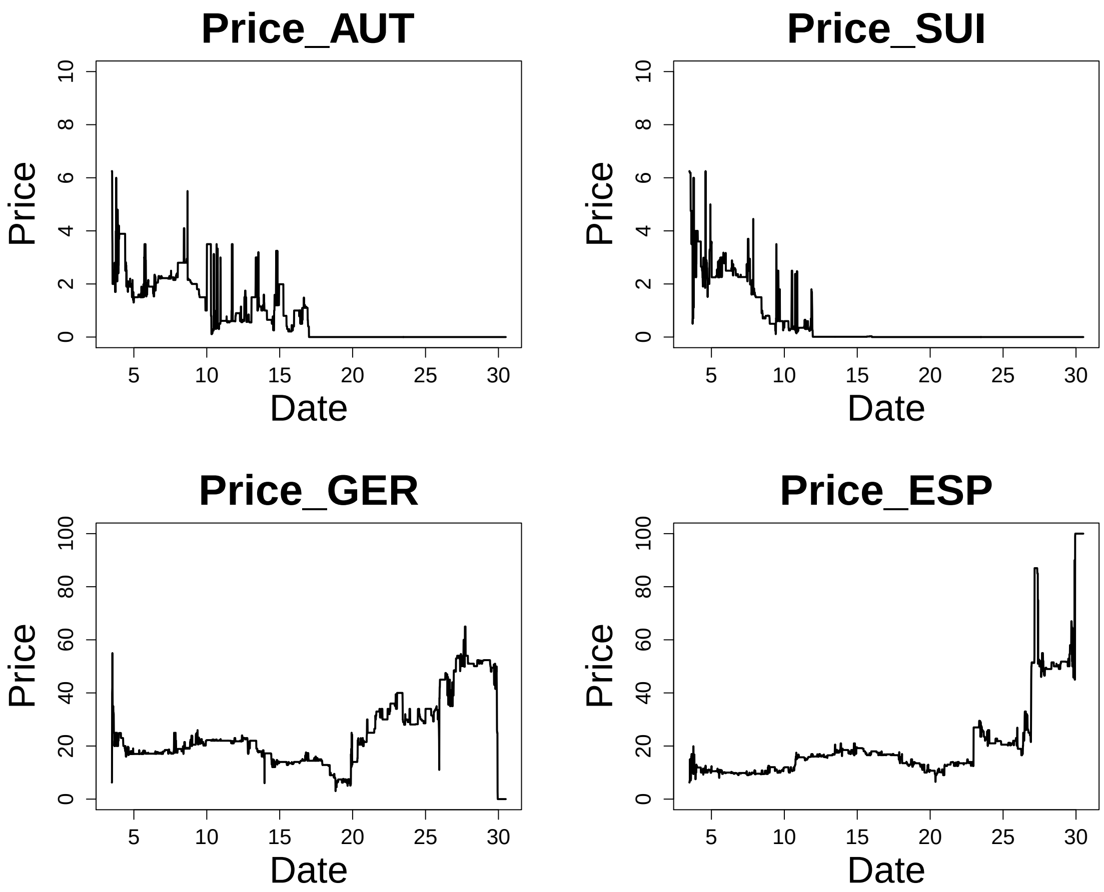


The measure of liquidity γ decays extremely rapidly over time for all team log prices.

Figure 2.11 shows an example of the γ for one of the log price series. The measure of illiquidity decays exponentially from day 3 (the beginning) to close to zero shortly after day 5 (2 days after inception) and continues close to zero for the remainder of the days up until day 30. We take this as further proof of the efficiency of the market, which can be observed for each individual price series under the γ measure. This provides further proof of Hypothesis 3.

Figure 2.12 shows price series for 4 selected teams from the 16 finalists; the two host nations, Austria (AUT) and Switzerland (SUI), both of whom did not succeed past the round-robin pool stage; and the two ultimate finalists, Germany(GER) and Spain (ESP).

Figure 2.12: Price Series of 4 Selected Tournament Teams 1



Note: The y-axis differs between the two teams in the top row versus the two teams on the bottom row, in order to demonstrate the fluctuations of the first two teams, whose prices fluctuate between zero and under 6.25, whereas the two teams on the bottom row show prices fluctuating between zero and 100.

2.6 Conclusion

In this chapter we analyzed the prediction market on the basis of the soccer European Championships 2008 as an example for an economic experimental market. According to the overwhelming proportion of published research, prediction markets provide excellent proxies for the aggregation of judgments for a variety of topics from sports to corporate business objectives, regardless of whether these experimental markets have been set up with real money or play money.

On our prediction market, we test whether prices are accurate forecasts of future events, such as victory or defeat in consecutive matches, and whether the market as a whole can be deemed efficient in the sense of rational expectations. We can answer that in our opinion, the soccer prediction market did an admirable job of acting as a focal point for public and private information, aggregated this information very effectively via the price mechanism and provided good predictions of the soccer games outcomes.

As the market was proven to be efficient in aggregate, the hypothesis that it can be used as a proxy for answering questions in real markets can be answered in the affirmative - as outlined above, if the prediction market exhibits price efficiency and fulfills the economic roles of a market, it can thus be considered a good proxy for a real financial market. In following chapters we aim to test whether behavioral biases are observed in individual market participants and how these can be dealt with from a risk management and early-warning perspective. Interestingly, despite potential individual biases and less-than-perfect predictions for individual teams, in aggregate the market functioned extremely well as an efficient predictor of future events. The Hayek hypothesis can be said to have been confirmed in this case.

Chapter 3

Overconfidence in Soccer

3.1 Introduction

3.1.1 Decisions, Preferences and Utility

Modern decision theory is based on the idea that probabilities of events happening can only be assessed properly with reference to future events (Hastie and Dawes, 2010). One of the three determinants of decision making is that the decision maker forms expectations concerning future events and outcomes flowing from each course of action. The expectations can be described in terms of degrees of belief or probabilities[5].

According to the subjective Bayesianist view, uncertainty, or subjective degrees of belief, are expressed in terms of probabilities. In probability theory, probabilities must satisfy certain axioms in order to qualify as being consistent. Agents' views of uncertainty are rational if and only if the probabilities assigned to uncertain events are consistent in the algebraic sense. According to Ramsey (1926), only logically consistent decisions can be rational decisions. DeFinetti (1937) shows that rational degrees of belief must conform to probability calculus and Kemeny (1955) makes the assertion that an agent's probability assignments are coherent, if the agent's subjective probabilities conform to probability calculus.

How people form their preferences is determined, as formalized by Savage (1954), by ordering preferences over uncertain options by assigning subjective probabilities to each

[5]The other two components of decision making are 1) There is more than one possible course of action under consideration, and 2) The consequences can be assessed on an evaluative continuum determined by current goals and personal values (see Hastie and Dawes, 2010).

option. Agents then update their subjective probabilities according to Bayes' Theorem (Bayes, 1763). As few people form preferences by maximizing monetary value, agent preferences over choice sets are measured by the utility assigned to each preference in the choice set. This approach goes back to work by the utilitarian school of Jeremy Bentham (1781), who proposed that maximization of overall (summed) utility would bring about *"... the greatest happiness of the greatest number"*, John Stuart Mill (1848), who gave rise to the term "homo oeconomicus" as the rational utility maximizer[6].

The expected utility model that is in use today was solved by Daniel Bernoulli as the St. Petersburg paradox, arguing that the paradox could be resolved if decision makers displayed risk aversion (Bernoulli, 1738). He suggested that people maximize expected utility rather than expected value. Expected utility theory was brought into the mainstream of twentieth century economics by John von Neumann and Oskar Morgenstern, who used the assumption of expected utility maximization in their formulation of game theory (von Neumann and Morgenstern, 1944). According to them, agents maximize expected utility when choosing among lottery outcomes.

Since Bernoulli, the theory of diminishing marginal utility, or risk aversion, has successively been formalized as the main method of systematically comparing preferences. Agents possesses a von Neumann-Morgenstern utility function if they rank uncertain payoffs according to (higher) expected value of their utility of the potential outcomes.

3.1.2 Expected Utility and Rationality

In order to resolve the problem that utility cannot be measured or observed directly, Samuelson (1938) developed the Theory of Revealed Preferences, which states that (in a perfectly competitive equilibrium) agents' behavior reveals the underlying relative utilities. These revealed preferences are revealed as prices; in equilibrium, the price reveals all relevant information that is in all of the traders' possession (see Hayek, 1945 or Grossman, 1976). Based on the work of Muth (1961), Lucas (1972) and Grossman (1981) formulated the theory of rational expectations - this assumes that agents' expectations may be individually wrong, but are correct on average (i.e. unbiased). These expectations use all relevant information in forming expectations of economic variables. Rational expectations do not differ systematically from equilibrium results, as it is assumed that the outcomes that are being forecast do not differ systematically

[6]Persky (1995) writes that J. S. Mill actually never used the exact term himself, but it was coined in response to his work.

from the market equilibrium results, i.e. predictions that deviate from perfect foresight are only random, but it is important to note that for efficiency to be present, it is not necessary for each agent to be individually rational, but rather the market is rational on average over all agents.

Rational expectations theory is the basis for the efficient market hypothesis EMH (Fama, 1965), which asserts that financial markets are informationally efficient. Under informational efficiency, market participants cannot consistently achieve returns in excess of average market returns on a risk-adjusted basis, given the information available at the time the investment is made. Depending upon which form of efficiency is assumed, weak, semi-strong or strong-form, all available information is assumed to be contained in the price of an asset. Samuelson (1965) applied the no-arbitrage condition to prove that correctly anticipated prices will conform to a random walk. This determines mathematically that today's price is the best predictor of tomorrow's price, i.e. that predictions do not add value and are futile, as all information is already contained in the price, as stated by the Efficient Market Hypothesis[7].

3.1.3 Alternatives to Rationality

There has been a great amount of theoretical and empirical research into the topic of rational expectations and the efficiency of markets in the twentieth century, establishing their precepts as current mainstream theory. Even though the term rationality is a fairly recent concept in economics - previous to that, economists assumed that people were motivated by self-interest (Arrow, 1986) - but ever since the mathematically overpowering concepts of von Neumann and Morgenstern (1944) and Savage (1954) took hold, it has dominated the theoretical world of economics and finance.

Even though Simon (1955) recognized that the two fields of economics and psychology were very far apart concerning explanations of learning and choice decisions, he *"... turned to the literature of psychology for the answer"* to the question concerning doubts about the "economic man" (Simon, 1955). Simon describes decision making as a search process guided by aspiration levels - a kind of benchmark that must be exceeded by an alternative for this to be a satisfactory decision alternative (Selten, 2001).

[7]That prices on stock markets evolve according to Brownian motion was actually first proposed by Louis Bachelier (1900), but not credited with the discovery until much later - mathematicians such as Kolmogorov were inspired by his findings from the 1930s onwards(see Courtault, 2000), but the first mainstream economists who realize what Bachelier had discovered were Savage and Samuelson in the 1950s (see Bernstein, 2005).

The individual finds decision alternatives iteratively in a search process. This Simon called "satisficing". In making their decisions, the rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make decisions. Simon argues that most people are only partly rational, and are emotional or irrational in the remaining part of their actions (Simon, 1957). Slovic and Lichtenstein (1971) came to the conclusion that people are not good *"intuitive statisticians"*, and Kahneman and Tversky (1972) found that *"[i]n his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all."*

Some of the strongest opposition to rational expectations theory comes from the field of psychology: Slovic (1969, 1972) analyzed common misconceptions of risk in decision-making and judgment, and recognized its relevance to finance. Tversky and Kahneman (1974) found that individuals do not act as a homo oeconomicus, but rather exhibit a number of heuristics and biases when making decisions under uncertainty, such as the representativeness, availability and anchoring biases. The experiments from a psychological viewpoint were formulated into prospect theory, an alternative theory to rational expectations, determining that people evaluate potential losses and gains in their decisions rather than an absolute level of wealth. Under prospect theory, individuals' utility functions are no longer exclusively convex (logarithmic) power functions, but are split into two segments around a reference point. Under this premise, individuals were found to be risk-seeking in the domain of losses, and risk-averse in the domain of gains (Kahneman and Tversky, 1979).

This beginning to serious theoretical competition to the EMH coincided with much empirical research raising doubts about classical finance theory, analyzing so-called anomalies that did not coincide with established theoretical precepts. Some of the earliest examples include Ball and Brown (1968) showing that abnormal returns were observable after earnings announcements by firms. Thaler (1987b) proved that the January effect (abnormally high stock market returns during the month of January) existed, and DeBondt and Thaler (1984) found that, in violation of Bayes' rule, most people tend to overreact to unexpected and dramatic news events.

Nevertheless, representatives of the established finance tradition have tried to negate the rising influence of behavioral finance, despite mounting empirical evidence - Fama (1998) published an article in 1998 rejecting that criticisms of the efficient markets hypothesis had any merit. Merton (1987) wrote in an article that the evidence against market efficiency was "premature", but since then, many papers researching biases in judgment and decision-making have been published. Judgment is the process by which opinions are formed, conclusions are reached, and critical evaluations are made. Individuals often make judgments spontaneously. Decision making is the process of choosing between alternatives and selecting available options (Gerrig and Zimbardo, 2010). Following in the footsteps of Kahneman and Tversky, researchers such as Hutchinson and Gigerenzer (2005) have suggested that humans have developed an “adaptive toolbox” (Gigerenzer and Selten, 1999) to find judgments that are usually correct using “fast and frugal heuristics” (Todd and Gigerenzer, 2001).

Even though researchers such as Gigerenzer (2008) or Hertwig et al. (2008) have found that heuristics can deliver correct judgments, often heuristics lead to errors. Many of these errors are systematic errors, judgments are biased. Judgments are said to be mediated by a heuristic when the assessment of a target attribute is substituted by heuristic attribute, and this systematically introduces a bias (Kahneman and Frederick, 2002). The representativeness bias and the availability bias were two of the first biases to attract attention by researchers (see Tversky and Kahneman, 1974). The online platform Wikipedia (2011) now currently lists more than 90 known cognitive biases. Two of the biases that have attracted much interest and published research are the overconfidence bias (see e.g. Svenson, 1981 or Odean, 1998b) and the disposition effect (see e.g. Shefrin and Statman, 1985).

3.1.4 The European Soccer Championships 2008

For 3 weeks in the summer of 2008, the world of soccer focused one single event: the finals of the European soccer championships[8], which is eagerly anticipated every 4 years. Not only fans from the respective countries of the 16 nations that had qualified, but according to UEFA (media release 76) at least 155 million TV viewers followed each of the 31 matches live. About 1 million lucky fans were able to obtain tickets in the preceding lottery. Those who had not been able to obtain tickets populated the fan zones that had been set up in the host cities, where large-screen TVs and expensive beer abounded, to watch the games live and demonstrate support for their team to the bitter end (at least for 15 of the 16 teams). Expectations varied amongst fans, depending on their differing levels of knowledge of and expertise on the teams.

Rivalry, home-team bias and wildly inflated expectations are all part of the process of defending one’s team regardless of objective facts. Quotes from coaches ranged

[8]The official designation was EURO 2008™

from *"Looking back at the match we deserved to lose. We were unable to play our combinations or play fluidly."* (German coach Joachim Löw quoted in Fisher, 2008) to *"We have become one of the great footballing nations. Our people can rejoice. And if our people are proud of us then we are proud of them."* (Turkey coach Fatih Terim quoted in Fisher, 2008). During an event such as this, emotions run high amongst players, coaches, fans and even politicians (*"As the prime minister I have to be balanced and collected. But last night I was speaking very differently about the whole thing, I wanted to kill."* Polish Prime Minister Donald Tusk quoted in Fisher, 2008).

Excessive confidence in one's own abilities on the part of the players/coaches as well as on the part of supporting fans was probably one of the most prevalent biases that could be observed. This meant that the European Championship provided a unique possibility to study confidence biases as a prime example of psychological dynamics.

Betting on sports has a long tradition, and expectations of increased gambling activity over the course of the Championship were high. British betting firm Ladbrokes estimated that more than $700 million would be wagered on matches during the three-week event (Scott, 2008). Betting on events such as this is highly sophisticated nowadays - there are increasing similarities between betting markets and the complexities and instruments of the financial markets. This led the author to establish a platform on which such financial transactions (or bets) could be monitored to observe potential excessive confidence behavior. Researching and understanding excessive confidence as a psychological dynamic in market situations has become increasingly important for behavioral economists and this paper aims to add to the existing body of knowledge on possible causes and effects of excessive confidence on trading activities, in particular on the existence of excessive confidence at the beginning or throughout trading actions.

3.1.5 Types of Overconfidence

The overconfidence bias has been formulated alternatively as (1) illusion of control (Langer, 1975), (2) the better-than-average effect (Svenson, 1981), or (3) miscalibration (Lichtenstein et al., 1982). Regarding (1), Langer (1975) identifies factors that are associated with skill situations such as practice, competition or choice. When these factors are introduced into chance situations, people then believe they are able to assert control over outcomes that are in fact due to chance. Presson and Benassi (1996) state that *"[p]eople often perceive more control than they actually have, notice covariation where none exists, and report inordinately high levels of prediction ability"*. Moore and

Healy (2007) study three measures of overconfidence. They identify a) overestimation (overestimation of one's actual ability, performance, level of control, or chance of success), b) overplacement (when people believe themselves to be better than others, such as when a majority of people rate themselves better than the median) and c) overprecision (excessive certainty regarding the accuracy of one's beliefs). Overplacement in Moore and Healy's terminology corresponds to Langer's better-than-average effect. Moore and Healy's categories overestimation and overprecision seem to the authors to be mainly contained within Langer's miscalibration category.

Regarding (3), evidence of miscalibration concerning confidence intervals or volatility estimates seems well established; people are notoriously ineffective at making such estimates correctly (see e.g. Lichtenstein et al., 1982). Odean (1998b) and others model overconfidence mainly as miscalibration, i.e. investors tend to underestimate the volatility of stock prices and forecast stock movements with too small confidence intervals.

3.1.6 Consequences of Overconfidence

Gervais and Odean (2001) find that success by skilled but overconfident traders will lead to higher trading volume and lower risk-adjusted performance for such a trader in the short run, but an increase in risk-adjusted performance over time. For less skilled traders who lose early on, there is no prediction of subsequent overconfidence or under-confidence, and such traders ought to fail at some point. *"Average levels of overconfidence are greatest in those who have been trading for a short time. With more experience, people develop better self-assessments. Since it it through success that traders become overconfident, successful traders, though not necessarily the most successful traders, are most overconfident"* (Gervais and Odean, 2001). Similar to overconfidence, Thaler and Johnson (1990) describe the house-money effect as an increase in risk taking when a trader has recent trading successes.

Behavioral finance studies of trading activity (e.g. Odean, 1999) have argued that (increased) trading can occur due to overconfidence, but recent studies (e.g. Grinblatt and Keloharju, 2009) however, question whether overconfidence leads to (increased) trading behavior. They argue that factors such as sensation seeking behavior are more prevalent in male subjects and thus cloud the issue of whether factors such as gender can be a correct predictor of overconfidence in trading. These examples indicate that overconfidence is often present in an investment or trading setting.

If overconfidence is inherent when investing, does it arise during the (repeated) investment activity as a form of reinforcement learning (for discussions of reinforcement learning in economics, see e.g. Camerer and Ho (1999), or Roth and Erev (1995)? We attempt to test for the presence of overconfidence in general in a trading setting and on the basis of these results to answer the question of nurture or nature - i.e. is overconfidence an acquired bias as posited by Gervais and Odean (2001) or is it present from the beginning as a character trait in analogy to the Grinblatt and Keloharju (2009) findings on sensation seeking?

Can trading strategies provide information on participants' over- or underconfident behavior? Classical investment literature (see e.g. Reilly and Brown, 2006) identifies investment strategies such as passive versus active strategies, and fundamental, contrarian or anomaly strategies within active management. As the trading market used for our analysis formed a zero-sum game in absolute money terms, passive strategies would not lead to success, but active strategies could focus on expectations in certain teams due to fundamental analysis or diversified strategies based on risk-reduction, for example.

An on-line prediction market linking participants' attitudes towards soccer and trading with actual trading behavior, along with objective ex-post outcomes of this behavior, allow us to address the following questions which are highly relevant for financial markets:

1. Is there evidence for overconfidence among the participants of our study?
2. Do overconfident participants trade more?
3. Can trading strategies be identified?
4. What are the characteristics of overconfident participants?
5. Are there connections between overconfidence and trading strategies?
6. Does overconfidence change over time?

As we intend to test for overconfidence in regard to trading and investment strategy application, we focus on the miscalibration bias, of the three measures referred to above. We do not intend to study overconfidence purely as regards the frequently used correct or incorrect choice of confidence levels, but rather focus on overconfidence as manifested by achieved trading results or strategies compared with initially documented intentions and expectations.

3.2 Method

An on-line electronic trading platform was established which allowed study participants to trade on the outcomes of soccer games during the European Soccer Championship 2008. The platform admitted students at two universities in Vienna (37.35% of the final sample of participants), as well as employees (27.35%), clients (15.0%) of an international consulting firm and other professionals (20.30%); see Table 3.1.

The broad professional and demographic distribution provided a realistic environment for trading stocks as one would in real-life markets. Supply and demand determine prices, investors have to deliberate the consequences of posting bid or offer trading indications and the price discovery mechanism condenses the varied information into one figure, the current price.

Before each participant was permitted to trade, a short psychological questionnaire was requested to be filled out online, as well as demographic data such as age, sex, occupation and residence.

3.2.1 Participants and Data

Participants

Of the 495 persons overall who entered the trading platform to place trades, 61 did not provide performance estimates or answer the psychological questionnaire. The 434 participants who provided the information required for study participation formed the initial study sample. Of all the participants in the study sample, 340 actively traded while 94 participants did not perform any trades, the 340 active participants formed the final study sample. Of this sample, most participants were male (63.24%) and between the ages of 20 and 29 (56.18%); see Table 3.1.

Data

The original data consisted of a file containing 50,047 individual lines of transactions; from these transactions, incorrect or partially missing data were cleansed, cancelled orders were removed, and all transactions from one participant were removed who had proved to have cheated and violated the general terms and conditions of the market. From each order containing a buyer and a seller, the price of the transaction was extracted and used to form each of the 16 teams' price series over time; this resulted in a file containing 48,757 individual transactions. These transactions were assigned to each participant in his or her portfolio at each instance in time.

Table 3.1: Demographic characteristics of the study sample (N=340)

Attribute	n	%
Sex		
Male	215	63.24%
Female	125	36.76%
Age		
< 20	15	4.41%
20 to 30	191	56.18%
30 to 10	74	21.76%
40 to 50	47	13.82%
> 50	13	3.83%
Profession		
Consultancy Employee	93	27.35%
Consultancy Client	51	15.00%
Student	127	37.35%
Other	69	20.30%
Total	340	100.00%

3.2.2 Research Design

Trading Platform Setup

For 2 Euros each, participants signed up to trade and were assigned 10,000 units of cash (football cents) and additionally a stock basket containing 100 units of each team's stock (e.g. Austria, Germany, Netherlands, etc.) also worth 10,000 cash units (the total of 20,000 cash units were equivalent to 2 Euros). Participants who signed up prior to the actual begin of the European Championships received 16 teams in a stock basket. Each stock started out at a price of 10,000 currency units / (16 teams * 100 shares) = 6.25 currency units per share, the price then fluctuating determined only by supply and demand.

The market was structured as a "winner-takes-all" market, meaning that the one team that ultimately won the championship would redeem at 100% (10,000 currency units), and all other teams at 0. This potential price fluctuation between 0 and 100 meant that the price at any given point in time could be interpreted as a probability of that particular team winning the trophy. Exactly at inception, each team was thus assigned a 6.25% chance of winning, which quickly changed following the initial trades (arbitrage was possible for a very short period at that time before prices settled down to an initial equilibrium).

This procedure has been effectively used by the Iowa Electronic Markets, one of the pioneers of on-line prediction markets (see e.g. Forsythe et al. (1992), or www.biz.uiowa.edu/iem). To make things easier, the consulting company sponsored the 2 Euros per participant, and the results were donated to charity after the event. Each participant signed an agreement that any profits would automatically be donated to charity, and that they would receive notification regarding any losses over and above the already paid-in initial 2 Euros to request the participant to donate the outstanding amount to the specified charity. This procedure eliminated any potential legal or excessive administrative efforts related to the collection of outstanding losses, relegating decisions on settlement of debts to the individual moral rather than the legal domain.

Trading Details

Those participants who signed up later during the course of the tournament, or purchased further stock baskets after individual teams had been knocked out of the tournament, received stock baskets comprising fewer than the initial 16 teams (the price of a stock basket remained constant at 10,000 currency units). The prices for each of the remaining teams adjusted accordingly. If, for example, 6 of the 16 teams had already been eliminated, each team's share price in a stock basket comprising the 10 surviving teams was thus 10 rather than the initial price of 6.25 when all teams were in the running. This corresponds to the fact that these surviving teams obviously also had a higher chance of advancing further towards the finals!

The on-line trading platform ensured that investors could trade at any moment in time 24 hours a day. The number of trades able to be placed was theoretically unlimited. Participants were able to buy additional cash units or stock baskets at any point in time on-line. Investors were able to see their total portfolio at any point in time and could make trades directly from the on-line interface to their portfolio.

The order book for each team was open, so each participant could see all bids and offers available for each team and trade correspondingly. Returns were calculated daily and posted on-line in a ranking table for all to see (trading returns were calculated as realized returns on a simple non-annualized percentage basis as there was only one trading period. Revaluations of portfolios with unrealized returns were not calculated).

In addition to the honor of having won, the most successful traders (those who achieved the highest cumulative return) were presented with prizes after the tournament had ended. The first in the category "Student" won a semester's tuition for free, the winner of the category "Consultant" won a wellness weekend, and the winner of the category "Client" won a trip to the country of the winning team. Both of these facts acted as powerful motivators for participants to trade, as evidenced by over 50,000 trades posted within the 3-week period; an additional motivator was the Activity Index indicating the cumulative number of matched trades for each investor along with his or her current ranking by activity.

Psychological Questionnaire

Before being allowed to trade, participants were first asked to register online and provide information on their demographic background and to complete a brief psychological questionnaire including a subjective prediction of the predicted return of their trading activities.

A psychological questionnaire asked participants to rate the importance they attached to soccer, the degree of expertise they believed to have on soccer, the degree of expertise they believed to have on making predictions about soccer games, and the influence they presumed luck to have on the outcome of the EURO soccer games. The following questions were posed (the questionnaire had a Cronbach's α of 0.736)[9]:

1. The importance of soccer ("compared to your other interests, what importance does soccer have for you?") was assessed on a six-point Likert scale from "very little importance" to "very high importance",

2. Supposed subjective expertise on soccer ("compared to others, how well versed are you in soccer") was measured on a six-point Likert scale from "much worse than others" to "much better than others",

[9] Classical test theory distinguishes three sources of variance: (a) true score variance, (b) error variance or measurement error, and (c) total scale variance, which is the sum of true score and error variance. Reliability is defined as the proportion of variance in an observed test score that is related to the true scores (McDonald, 1999). Cronbach's alpha (or coefficient alpha), see Cronbach (1951), is one of the most popular reliability measures of internal consistency of tests (Miller, 1995).George and Mallery (2003) provide the following rules of thumb for the acceptability of Cronbach's α : $> .9$ – Excellent, $> .8$ – Good, $> .7$ – Acceptable, $> .6$ – Questionable, $> .5$ – Poor, and $< .5$ – Unacceptable.

3. Supposed subjective expertise on predicting the outcome of soccer games (”compared to others, how well can you predict the outcome of soccer games” was measured on a six-point Likert scale from ”much worse than others” to ”much better than others”, and

4. The supposed influence of luck on soccer game outcomes (”according to your opinion, which influence does luck have on the outcomes of the games of EURO 2008”) was measured on a six-point Likert scale from ”very little influence” to ”very high influence”.

5. The questionnaire asked participants to give a numeric estimate (in percent) of how their capital at the beginning would develop over the course of their participation. Participants were informed that, for example, ”+50%” means that they win an additional 50% of their seed capital, ”0%” means that their capital at the end will be as high as their seed capital, and that ”-50%” means that they will lose 50% of their seed money. Instructions on the impact of trading on returns was posted on-line in the FAQ section of the platform.

We attempt to localize overconfidence within trading patterns as observed ex-post versus the ex-ante indications, as well as being able to test for different aspects that might lead to overconfident behavior. The answers to the psychological questionnaire formed one part of the evaluation of potential overconfidence. The results of participants' intentions as measured by the questionnaire must be put into context by combining intentions with achieved results. A first comparison of intentions with trading results can be derived from comparing expected returns from the questionnaire with actually achieved returns. The estimates of expected returns and the comparison between expected returns and realized returns allow examining possible overconfidence manifested as miscalibration among the study sample. Further aspects of the results of participants' trading behavior can be measured using traditional portfolio management parameters, such as risk, diversification and related measures.

3.2.3 Measures

On the online platform, all participants were ranked in terms of absolute return, i.e. the percentage of their starting capital that had been won or lost since the participant started trading. The return calculations were performed daily and rankings updated online accordingly. Only realized returns were considered in the online calculations and rankings, i.e. an increase in value of a stock in a participant's portfolio only increased the published portfolio value after the stock had been sold and the profit transformed from a paper (unrealized) gain (or loss) to a realized profit (or loss). As the online rankings were based on absolute realized returns, we also base our return calculations on this premise.

Realized returns are only a part of the measurement of investment success or failure. It is important to measure on the basis of which risk these returns were achieved. We use the standard measure of risk from portfolio management, the standard deviation of returns to measure the fluctuations of by-trade returns around their mean. In real-world portfolio management, revaluation and performance results from a portfolio are usually calculated at the end of each business day. In contrast, we perform calculations of all parameters such as returns, risk, etc. on a trade-by-trade basis, not at a cutoff point (such as at the end of a day), as there is no good cutoff point that stays constant – games were held at differing times on differing days of the week. We compare the ex-ante data together with the ex-post results from trading activities such as:

1. Absolute number of trades (activity)
2. Relative number of trades per actual day of trading (daily activity - controls for late entrants)
3. realized return
4. Riskiness of traded stock (measured as standard deviation of returns)
5. Diversification of stock holdings as a risk-reduction measure

The number of trades is measured on a total basis as well as on a per-day basis in order to determine whether trading activity is related to overconfidence. Other authors such as Odean (1999) have found a connection between trading activity and overconfidence. In portfolio management, the question of how to allocate one's portfolio of stocks (i.e. how many different stocks, and in which proportion of the total portfolio) is of paramount importance, as the allocation substantially affects both risk and return results (see Markowitz, 1952). In order to determine participants' allocation strategies, we derive a diversification measure to determine whether asset allocation (and connected to this, risk-return combinations) can be related to overconfidence.

Determination of an Overconfidence Index

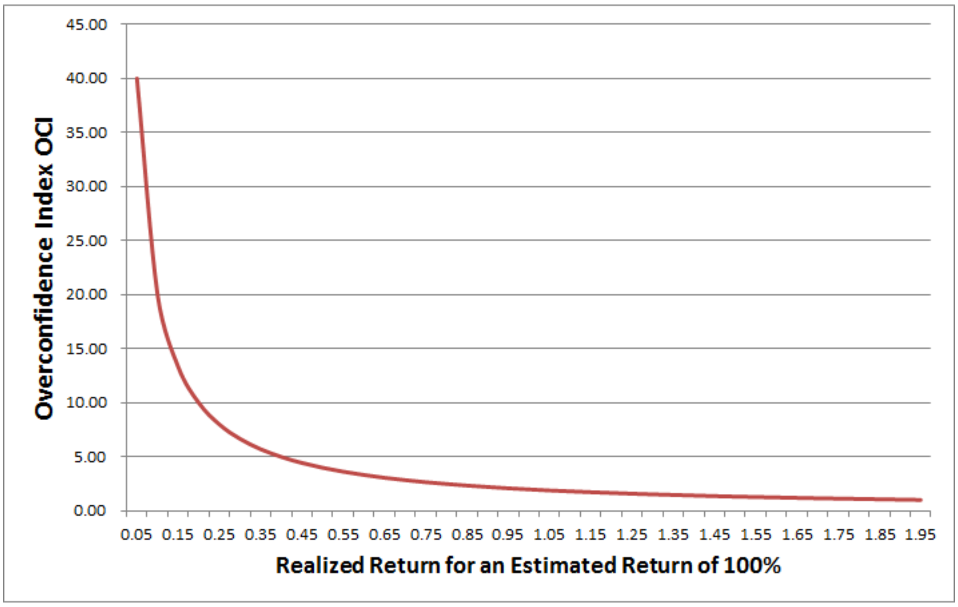
A first impression of the existence of overconfidence starts with a comparison of realized returns versus forecast returns. If a participant forecasts a positive return (profit) but achieved a negative return (loss), the participant can be viewed as having been overconfident in his or her estimations. This falls under the categoriziation of overconfidence as expressed by miscalibration. A more precise measure of the calibration of participants' return estimates rather than simply looking at the positive or negative sign of return figures can be calculated via an Overconfidence Index (OCI). We define the OCI as

$$OCI \equiv \frac{FV^{t_0}_{expected}}{FV^{t_n}_{realized}} \tag{3.1}$$

where $FV^{t_0}_{expected}$ is the expected future value of the portfolio at time t_0 (the beginning of a participant's trading) and $FV^{t_n}_{realized}$ at time t_n is the realized futures value at maturity. The OCI has a value of 1 when the realized return equals the estimated return, indicating a well-calibrated investor. OCI values > 0 but < 1 indicate underconfidence: The OCI converges toward 0 the higher the realized return is vs. the expected return, i.e., the higher the degree of underconfidence. OCI values > 1 indicate overconfidence: the OCI increases the less the realized return is relative to the estimated return, i.e., the higher the degree of overconfidence. Our overconfidence index OCI shows every level of overconfidence regardless of the sign of the achieved versus estimated absolute returns (i.e. in percentage terms). The OCI changes with both the relationship between achieved and estimated returns as well as with the absolute level of these returns between -100% and $+\infty$.

Figure 3.1 shows the progression of the overconfidence index, plotting potential realized returns in relative terms (not in percent) to the estimated return, against the corresponding OCI for a given return. For instance, a participant who expects to double the initial capital over the lifetime of the market has a return estimate ε of 100%. The graph indicates the participant's OCI for realized returns π between -95% (0.05 in relative terms) and +95% (1.95 in relative terms) depending on the actually realized return (e.g. an estimated return of -95% in relative terms is 1 + (-95 / 100) = .05; and +95% is 1 + (95 / 100) = 1.95. If a participant estimated a return of +100%, i.e. a relative return of 1 + (100 / 100) = 2, and actually achieved a return of -95%, i.e. .05, then his OCI would be 2 / .05 = 40, indicating extreme overconfidence).

Figure 3.1: Overconfidence Index vs Realized Return for Different Levels of Estimated Returns



Determination of Diversification

Portfolio diversification, or the concept of "don't put all your eggs into one basket", is based on the recognition, formalized by Markowitz (1952), that spreading the assets in a portfolio over different assets or asset classes reduces risk. The result of this is that the amount of expected return per unit of risk in increased by shifting invesmtents from only one asset to two or more assets in a portfolio. An investor who puts all his eggs into one basket, i.e. invests the total portfolio value into only one asset, can be described as having a concentrated or focused strategy.

In contrast, another investor who spreads her portfolio over several different assets is following a diversified strategy. The diversified strategy is, by definition, following a less risky course than a strategy focused on a speculation on a single asset. The nature of diversification implies that by diversifying, the risk of experiencing an adverse situation such as total loss is substantially lowered. This, however, comes at the cost of not preserving the possibility of an extremely good outcome. For example, if a participant had immediately sold all other teams' stock and invested everything in Spain stock (the final winner of the tournament), this participant would have followed the riskiest possible course - either winning big at the end when Spain won, or losing everything. The potential for such a gain is lessened by diversifying into other stocks (the teams who obviously did not win the tournament), but by judiciously trading during the course of the tournament, this participant would not have to fear losing everything, but would be able to achieve reasonable returns, resulting in an advantageous risk-return relationship (for a discussion of the implications of diversified versus concentrated investment strategies, see Borge, 2001).

In portfolio management and performance measurement, indicators for measuring the extent of the diversification of a portfolio are not usual, so we resort to designing a custom-made measure for our purposes. This measure needs to fulfil different restrictions: it needs to be comparable over all participants, so it must be a relative measure. Thus, a standardization to values between zero and one seem sensible. Also, it must accommodate a differing maximum number of outstanding teams, i.e. maximum possible diversification in a way that it is still able to compare the level of diversification when 16 teams are in the tournament, just as when fewer teams form the maximum number.

To examine the connection between overconfidence and portfolio diversification, a comprehensive measure of portfolio diversification is established which considers possible changes of portfolio diversification over time. This measure also takes into account the inherent difference between bets placed in the course of a knockout championship and the stock market. In both settings, participants may follow investing/betting strategies which range from diversified to focused. However, unlike on the stock market, the number of stocks that were available for study participants to choose from, narrowed systematically over the duration of the study - from 16 teams in the beginning of the championship to 2 teams before the final, implying that a fully diversified strategy would comprise 16 different stocks in the portfolio at the beginning, but only two different stocks by the end.

Thus, in order to determine a participant's degree of portfolio diversification at any given point in time, the number of outstanding teams has to be considered. This is achieved by deriving a variant of the Herfindahl-Hirschman Index (HHI), a measure of concentration from the family of entropy measures (Jacquemin and Berry, 1979). Following the suggestion of Adelman, quoted in Kelly (1981) we used the reciprocal of the HHI, adding an additional parameter n to include the number of outstanding stocks. We call this measure of diversification of portfolios the "Degree of Diversification" (DoD); it is new to the best of the authors' knowledge. This measure is defined as follows:

$$DoD \equiv \frac{1}{n[\sum_{i=1}^{n}(\omega_i^2)]} \tag{3.2}$$

where n is the number of outstanding stocks (soccer teams), n $\in$ [0, ..., 16] and ω is the individual stock weight calculated as the number of shares per team in the portfolio divided by the total number of shares in the portfolio.

The DoD can assume values of between 0 and 1. A DoD of 1 indicates total diversification, e.g. when there are 16 teams at the beginning of the competition and a participant holds an equal number of shares in each of the 16 teams, or when there are two teams at the end of the competition and the participant holds an equal number of shares in each of the 2 teams. The DoD becomes smaller and tends towards 0 the smaller the percentage of teams in which stocks are held at any given point of time and the more unequal the distribution of stock across the teams in which stocks are held, e.g. when there are 16 teams outstanding in the competition and stocks of two teams are held in the portfolio with equal weight in each, this results in a DoD of 0.125, calculated as 1 / 16 x (0.5^2 + 0.5^2). When there are two teams outstanding in the competition and the participant holds 90 shares of team A and 10 shares of team B, this results in a DoD of 0.61, calculated as 1 / 2 x (0.90^2 + 0.10^2).

The DoD expresses a participant's degree of diversification or concentration among the outstanding teams in the competition at any given point in time of the competition. Because a participant's DoD changes in the course of the competition, the mean and median DoD values over time are used to indicate a participant's general level of diversification throughout the competition (the mean DoD is used as an expected value, and the median DoD to control for potential outliers).

3.3 Results and Discussion

3.3.1 Subjective Return Estimates

Participants' expectations about how their initial capital would develop until the end of their participation ranged from -100% (i.e., to lose everything invested) to +1000% (i.e., to increase their initial capital 10-fold). Most of the return estimates were between -100% to +100% with outliers of up to +1000%. The summary statistics are shown in Table 3.2 for both expected and realized returns. We group participants' expectations and results into 5 categories: a) Those who lost virtually all their capital (< -99%), b) those who lost money but did not go totally bankrupt, i.e. realized returns > -99% and < 0%, c) those who expected not to make or lose money (none actually achieved that result), d) those who achieved postive but not huge returns, and finally e) those who achieved huge returns (defined as greater than 100%).

Table 3.2: Estimated Returns

Expectation/Achievement	Expected No.	Expected %	Realized No.	Realized %
Lose virtually everything (< -99%)	20	5.9%	83	24.4%
Achieve negative return but not all (> -99% to < 0%)	56	16.5%	126	37.1%
No return (= 0%)	72	21.2%	0	0.0%
Achieve positive but not huge return (> 0% to 100%)	174	51.2%	80	23.5%
Achieve huge return (> 100%)	18	5.2%	51	15.0%
Sum	340	100.0%	340	100.0%

Comparing the results from Table 3.2, it is obvious that expectations and achievements diverge considerably. The first two rows of Table 3.2 show that considerably more investors achieved below-average results than they expected; only 22.4% expected to lose money, whereas fully 61.5% actually made a loss, almost three times the expected number (see the first two rows of Table 3.2). In other terms, 77.6% of the participants were expecting not to lose money, but only 38.5% were able to fulfil that goal. This relationship shows a dichotomy of expected returns versus realized returns.

But overconfidence can occur at all levels (excepting those 6% who estimated returns of around -100%) – someone who forecasts a return of e.g. 0% can have lost money, which still makes this particular participant overconfident. Table 3.3 groups participants according to whether estimated and realized returns were positive, neutral, or negative.

Table 3.3: Overconfident, Underconfident and Well-Calibrated Trading (n=340)

Realized/Estimated	Estimated < 0	Estimated = 0	Estimated > 0
Realized < 0	Well-calibrated Losers 15% (n= 51)	Overconfident Losers 12.6% (n = 43)	Overconfident Losers 33.8% (n = 115)
Realized > 0	Underconfident Winners 7.4% (n = 25)	Underconfident Winners 8.5% (n = 29)	Well-calibrated Winners 22.6% (n = 77)

The numbers in Table 3.3 represent a first indication of whether participants were overconfident or underconfident in their initial return estimates as compared to their realized returns; it shows participants' trading results versus their expectations categorized as overconfident (expectations exceed results) and underconfident (results exceed expectations) on an absolute scale, i.e. negative return or positive return (no participants achieved exactly 0% return). The distribution of returns shows that 37.6% of all participants are well-calibrated (i.e., participants whose positive, zero, or negative return estimates matched whether their realized returns were positive, zero, or negative). An additional 46.4% of participants are overconfident, and 15.9% underconfident.

These results are a first rough indication of overconfidence in our sample considering only mis-estimation over the three categories a) negative return, b) no return or c) positive return. It does not consider those who are over- or underconfident within a respective category. A participant could have forecast a large positive return but only achieved a much smaller, but still positive return (e.g. someone who estimated +50% and only achieved +25%). Conversely, the same mis-estimation could be concealed within the negative return category. A more refined analysis is required to find potentially over- or underconfident traders hidden within the groups termed well-calibrated winners and well-calibrated losers.

3.3.2 OCI Results: Is there Evidence of Overconfidence?

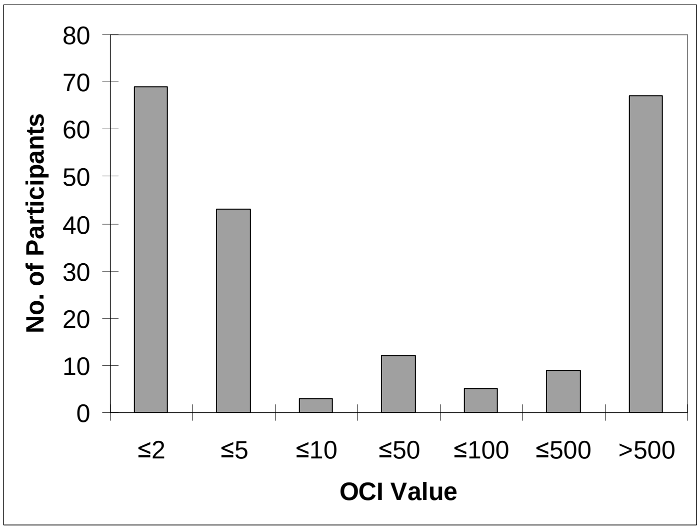
Table 3.4 shows that the initial results can be confirmed by a more detailed look at participants' Overconfidence Index results. Of the 340 active participants, 208 were found to be overconfident by comparing their return predictions with actual achievements; these participants have an OCI of greater than one.

Table 3.4: Evidence of Overconfidence

Over/Underconfidence	OCI Value	Number	% of Overconfident	% of Total
Underconfident	< 1	131		38.53%
Well Calibrated	$= 1$	1		0.003%
Slightly Overconfident	1 - 2	69	33.17%	20.29%
Moderately Overconfident	2 - 5	43	20.67%	12.65%
Highly Overconfident	5 - 500	29	13.94%	8.53%
Extemely Overconfident	> 500	67	32.21%	19.71%
Sum Overconfident		208	100.00%	61.18%
Sum Total		340		100.00%

Of the 340 active participants who actively traded, 208 (61.2%) have an OCI > 1, 131 (38.5%) are underconfident with an OCI < 1, and 1 participant exactly realized the predicted return. This result confirms the overall existence of overconfidence within our sample ($t(339) = 4.596$, $p < 0.001$, two-tailed).

Figure 3.2: Distribution of Overconfidence



Note: The Figure shows the distribution of absolute overconfidence values. It is notable that many participants are slightly overconfident, showing an OCI of between 1 and 2; and many participants are extremely overconfident, showing an OCI of greater than 500.

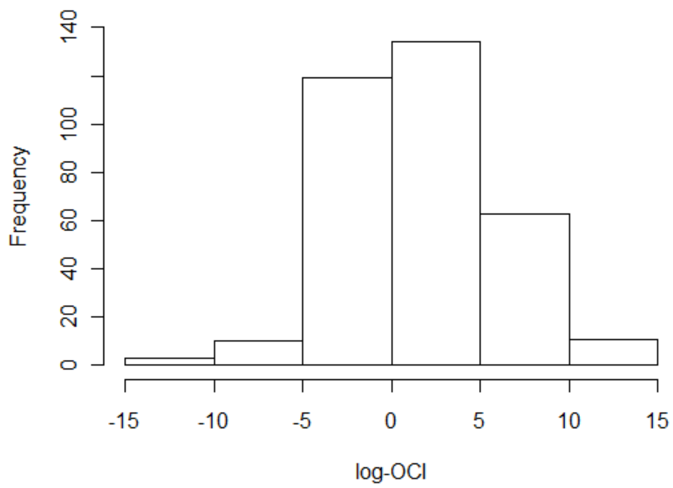
The resulting frequency distribution of the 208 overconfident participants is shown in Figure 3.2. Of these 208 traders, many participants (33.17%) are slightly overconfident (OCI = 2), 20.67% are moderately overconfident (OCI = 5). Quite a large number of participants (32.21%) are extremely overconfident (OCI $>$ 500). The first research question can be answered in the affirmative, there is ample evidence of overconfidence among participants in the study.

3.3.3 Who are the Overconfident Participants?

What then is the difference between overconfident and underconfident investors? Figure 3.3 shows the number of participants grouped in levels of log-confidence. One immediate observation is that the distribution of over- or underconfidence is heavily right-skewed which is confirmed by a D'Agostino test for skewness (skew = 0.51, z = 2.46, p-value = 0.006).

Table 3.5 contrasts the characteristics of different groupings of participants grouped by their level of over- or underconfidence. Median values are shown for results whose distributions are highly skewed (realized returns and volatility of these returns). The median values are grouped much closer than the means, confirming the t-test results which show no significant differences between the over- and underconfident groups.

Figure 3.3: Histogram of log Confidence



The first comparison is between the group of overconfident versus the underconfident participants, shown in the left- and right-most columns of Table 3.5. Most values are not significantly different, the obvious exceptions being the result values (log-OCI, expected return and realized return). This fact is interesting in itself, as overconfidence is not obviously related to items such as sex, age, activity (this value is even opposite in sign to expectations), or portfolio diversification. However, it is of interest that the importance of the questionnaire answers is highly significant (concerning importance of the subject matter) and at least marginally significant (p = .0628) for participants' prediction ability estimates. The factors will be analyzed in more detail in the following chapter.

Furthermore, analysis of the group of most overconfident participants (identified by a log-OCI value of greater than 5) versus all overconfident participants (the first and second columns in Table 3.5 shows that the only predictors (versus outcome variables such as realized return) where extreme overconfidence differs from general overconfidence is that the responses on the questionnaire answers are all slightly higher for the very overconfident group versus all overconfident participants which in turn are higher than the values for underconfident participants (even though regression analysis does not show a formalized continuous increase that is significant).

In summary there seem not to be any easily discernible identifiers for over- or underconfidence judged just by differences in statistical data based on trading results. In order to find common characteristics of overconfident behavior we analyze the over- and underconfident groupings in more detail.

Table 3.5: Summary Statistics by Confidence Category

Coefficient	All OC Participants	Very OC Participants (log OCI > 5)	All UC or WC Participants	t_{diff}/χ^2 Value	p-Value
Number of Male	134	53	81	0.21^a	0.649
Percent Male	64.4%	71.6%	61.4%		
Number of Female	74	21	51		
Percent Female	35.6%	28.4%	38.6%		
Category Participant Number	208	74	132		
Mean Age	30.67	33.00	29.98	0.67	0.499
Mean Answer: Importance	2.98	2.97	2.55	2.66***	0.008
Mean Answer: Knowledge	3.13	3.20	2.92	1.36	0.174
Mean Answer: Prediction	2.96	2.97	2.55	1.86*	0.062
Mean Answer: Luck	3.63	3.69	3.57	0.48	0.625
Mean Return Estimate	42.47%	25.55%	−13.56%	7.12***	< 0.001
Median realized return	−70.34%	−99.95%	58.16%	−7.79***	< 0.001
Mean Activity	100.12	149.19	175.89	−2.51*	0.012
Mean log OCI	3.61	8.29	−1.55	15.01***	< 0.001
Mean Degree of Diversification	0.55	0.40	0.53	0.49	0.621
Median Stdev of Returns	0.11	0.14	0.13		

Note: OC: Overconfident, UC: Underconfident, WC: Well-Calibrated

All tests are Welch's two-sample t-tests except those marked ^{a}, which are two-sided χ^2 tests for proportions. The t-test p-Values indicate that the value for the UC/WC category is significantly different from the OC category at the given level. The p-Value in the first row indicates that, from a χ^2 two-sample test for equality of proportions, the null hypothesis cannot be rejected that the fraction of men who are overconfident rather than well-calibrated or underconfident is equal to the fraction of women who are overconfident, i.e. men are equally overconfident as women within the sample.

*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

3.3.4 Overconfidence and Trading Frequency: Do Overconfident Traders trade more?

Previous research (e.g. Odean, 1999) finds that trading often leads to overconfidence: if results confirm success, then Odean finds that traders become overconfident, ultimately losing out by reducing their gains due to too frequent trading and additionally incurring substantial transaction costs. We analyze whether we can find evidence of a connection between trading activity and overconfidence as expressed by the OCI.

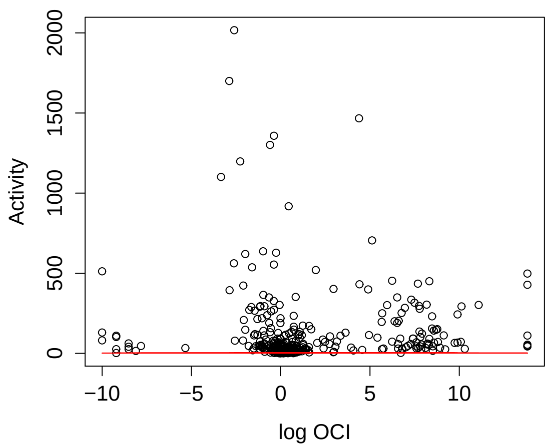
In order to determine whether overconfidence influences trading activity, the log-OCI is regressed on the absolute number of trades performed by each participant (controlling for those who did not trade at all):

$$logOCI_i = \alpha + \beta A_i + \epsilon_i \tag{3.3}$$

where i are the 340 active participants, A denotes the number of trades per participant (activity). The results are shown in Figure 3.4.

The results of the regression do not show a connection between overconfidence and the number of trades executed. An F-test shows that neither the intercept nor the gradient of the regression line are significantly different from zero: F-statistic: 0.03692 on 1 and 338 DF, p-value: 0.8477. Previous research by Odean (1999) finds a positive relationship and derives the hypothesis that overconfidence leads to increased trading activity, but we cannot confirm this hypothesis from our data.

Figure 3.4: Relationship between Trading Activity and Overconfidence



3.3.5 Demographic Differences in Overconfidence

Overconfidence and Sex

The analysis of participants' OCI by sex (see Table 3.6) shows that 62.3% of men and 60% of women were overconfident. A χ^2-test for categorical proportions shows that both of these proportions of overconfident men or women are significantly different from a theoretical mean of 50% ($\chi^2(2) = 17.18$, $p < .001$), proving that significantly more men and women than expected by chance alone were overconfident. The percentage of overconfident men is not significantly different from the percentage of overconfident women (see Table 3.6), so the assumption that men tend to be more overconfident than women in a trading setting as has been found by other authors such as Barber and Odean (2001) cannot be confirmed in our sample.

While the groups of men and women do not show statistically significant difference in OCI values, men tended to be more extreme in their views than women regarding their return expectations - the mean as well as the median return expectations shown in Table 3.6 are significantly different from each other. Men also traded significantly more actively than women did with a mean number of trades of 159.63 for men and 77.77 for women.

Men and women do not show significantly different attitudes toward risk in the mean (neither the standard deviations of returns are significantly different from each other, nor the median values), but, male risk-taking has much more variance than female risk behavior (the mean differs substantially from the median for men but not so much for women). Interestingly, diversification among our sample proves to be a female attitude rather than a male attitude (the degree of diversification is significantly lower for men in the sample).

Overconfidence and Age

A correlation between age and overconfidence (OCI) does not show any significant connection, with a value of $\rho =$ -0.02). This may have to do with the fact that the goal of all participants was to achieve a maximum return within a given (fairly short) investment horizon, mainly using house money[10]. Thus most of the usually observed differences

[10] House money is a concept developed by Thaler and Johnson (1990) and describes the phenomenon of increased risk seeking in the presence of a prior gain. They quote as an example people who win in the casino and display substantially increased risk-seeking behavior with the additionally earned

Table 3.6: Differences by Sex

Category	Men	Women	t/χ^2/W-Value	p-Value
Overconfident	62.30%	60.00%	0.09	0.75[a]
Mean Return Expectation	28.82%	6.78%	2.68***	0.007[b]
Median Return Expectation	20.00%	0.00%	4095.5**	0.036[c]
Mean Return Achieved	49.38%	−10.15%	2.94***	0.003[b]
Median Return Achieved	−34.71%	−16.50%	5633.0	0.104[c]
Mean No. of Trades	159.63	77.77	3.74***	< 0.001[b]
Median No. of Trades	54.00	37.00	4180.0*	0.061[c]
Mean Risk (Std.dev.)	127.03%	48.36%	0.67	0.49[b]
Median Risk (Std.dev.)	11.53%	12.66%	4604.0	0.395[c]
Mean Diversification	51.95%	59.91%	−3.85***	< 0.001[b]
Median Diversification	50.53%	59.41%	6170.0***	0.003[c]

Note: [a] denotes χ^2-tests for categorical proportions, [b] denotes Welch's two-sample t-tests and [c] denotes Mann-Whitney/Wilcoxon rank-sum tests. The p-Value in the first row indicates that, from a χ^2 two-sample test for equality of proportions, the null hypothesis cannot be rejected that the fraction of overconfident men is equal to the fraction of women who are overconfident.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

in investment behavior between age groups might not have been relevant within the specialized framework of our experiment. In contrast to expectations, the youngest member was slightly underconfident, and the oldest member was vastly overconfident!

3.3.6 Psychological Influences on Overconfidence

The question we ask is: Does the subjective importance of soccer and perceived influence of luck influence overconfidence?

The Importance of Soccer

Question number 1 asked participants how important soccer is for them. We hypothesize that the higher the relative importance for a participant, the higher

a) the participant's subjective expertise,

b) subjective comparative prediction expertise,

c) subjective performance prediction and

d) overconfidence index, and

money than with the money they brought with them to the casino, giving rise to the term "house money" as a result.

e) the lower the estimated luck influence should be.

The gender split: As might be expected, soccer was deemed of medium or high importance by 70% of all men, whereas 65% of all women expressed low intererest in the topic, as shown in Table 3.7.

Table 3.7: Soccer Importance by Gender

Gender	Low Importance	Medium Importance	High Importance	Sum
Male	29.8%	51.6%	18.6%	100.0%
Female	65.6%	28.0%	6.4%	100.0%

Table 3.7 shows the distribution of professed interest in soccer by gender. Unsurprisingly, interest in soccer seems to a be predominantly male pastime. Nevertheless, Table 3.8 shows the outcome of the hypotheses relating to the importance of socce for the participants over all evaluated participants (n=340) in the column "Soccer Importance".

Table 3.8: Questionnaire Answers Correlation Matrix

Category	Soccer Imp.	Subj. Expert.	Pred. Expert.	Luck Influence	Perf. Forec.	log-OCI
Soccer Imp.	1	0.80***	0.60***	0.06	0.12	0.10
Subj. Expert.	0.80***	1	0.73***	0.01	0.13	0.09
Pred. Expert.	0.60***	0.73***	1	0.06	0.18***	0.14
Luck Influence	0.06	0.01	0.06	1	−0.03	0.06
Perf. Forec.	0.12	0.13	0.18***	−0.03	1	0.26***
log-OCI	0.10	0.09	0.14	0.06	0.26***	1

Note: Soccer Imp. is shortened from Soccer Importance, Subj. Expert. from Subjective Expertise, Pred. Expert. from Prediction Expertise, Perf. Forec. from Performance Forecast.
*** denotes significance at the 1% level.

Table 3.8 shows the outcome of the correlations of soccer importance and the other psychological questionnaire items together with the overconfidence index OCI. It can be seen from the table that the answers to the questions whether soccer is of importance to participants, and whether participants believe they have expertise in soccer and in their ability to predict match outcomes are highly significantly correlated.

Correlation tests for Pearson correlation for these items (performed with cor.test in R) show the following results:

- Test for correlation of Soccer Importance with Subjective Expertise: $t(338) = 24.20$, p-Value < 0.001

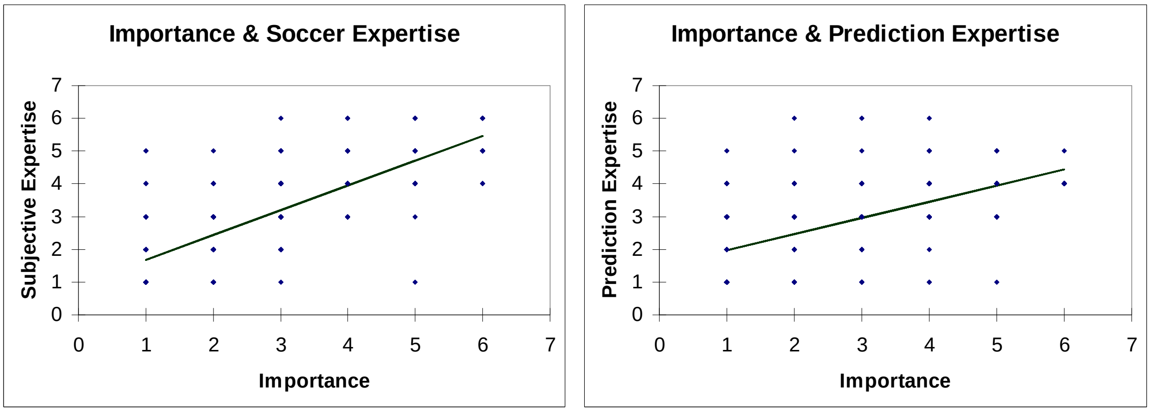
- Test for correlation of Soccer Importance with Prediction Expertise: $t(338) = 13.85$, p-Value < 0.001
- Test for correlation of Performance Forecast with Prediction Expertise: $t(338) = 3.42$, p-Value < 0.001
- Test for correlation of Performance Forecast with log-OCI: $t(338) = 5.02$, p-Value < 0.001

None of the other categories show highly significant correlations. Only the tests that proved high significance of the correlation values are shown above.

- **Hypothesis 1:** Participants for whom the subject matter (soccer) is important will also profess to having above-average expertise of the subject matter.
- **Result:** This can be confirmed (compare the left graph of Figure 3.5). A correlation (ρ) of 0.80 in Table 3.8 shows that those participants who are interested in the subject matter also regard themselves to be knowledgeable on the subject.
- **Hypothesis 2:** Participants for whom the subject matter is important will also profess to having above-average expertise in predicting the outcomes of games.
- **Result:** Participants are not quite so certain of their expertise in predicting the outcome of soccer games ($\rho = 0.60$, see Table 3.8) as they were of their knowledge, but there is still a highly significant positive correlation between a participant's intererest in the subject matter and his/her confidence or lack thereof in being able to forecast the outcomes of games, depending on his/her interest in soccer (compare right graph in Figure 3.5). One might expect to see negative correlation between subjective expertise in predicting the outcome of games and the last question – how important is the factor luck in the outcome of games. However, there was practically no correlation between the answers to these two questions ($\rho = 0.06$, see Table 3.8).

The graphs in Figure 3.5 show regressions between between professed soccer importance and estimated soccer expertise / prediction expertise. The answers in the questionnaire show that those participants who are interested in soccer also regard themselves to be knowledgeable on the subject (left graph). The right-hand graph shows that the interested participants are slightly less confident of their general prediction abilities, but there is still a fairly strong correlation between interest in soccer and professed prediction expertise.

Figure 3.5: Importance and Soccer / Prediction Expertise



Note: The two Figures above show regressions of different answers to the questionnaire items, on the left, participants' answers to how good their soccer expertise is versus the importance of the subject, and on the right is the relationship between participants' self-reported expertise in prediction versus the importance of the subject to them.

- **Hypothesis 3:** Participants for whom the subject matter is important will also have higher forecasts of their performance in trading on this subject.

- **Result:** This hypothesis could not be confirmed, the correlation $\rho = 0.12$ is not significant, see Table 3.8). Obviously participants neither consider superior subject matter expertise a guarantee for good trading results, nor a lack of expertise to preclude them from achieving good results.

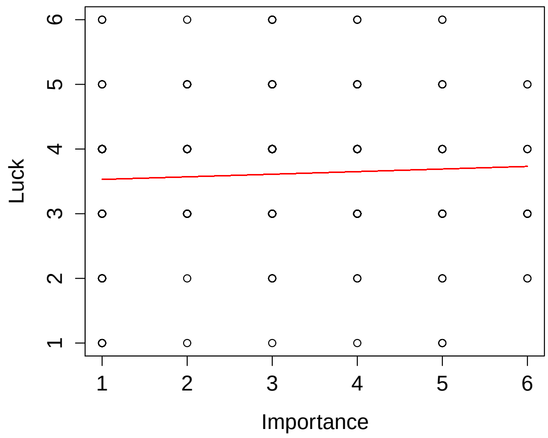
- **Hypothesis 4:** Participants for whom the subject matter is important will also rate the influence of luck on the outcome of games lower.

- **Result:** A negative correlation between these parameters could not be confirmed from the data ($\rho = 0.06$ is not significant, see Table 3.8). We would have expected participants who are interested in the subject and who also consider themselves knowledgeable about the subject to differ in their perception of luck on the outcome of soccer games.

In order to pursue this analysis further, we split the participants answers into three groups (low, medium and high knowledge) and ran the analysis again. One might expect those with little knowledge to rate the outcome as determined by luck, or even those with high knowledge to "know" that luck determines the outcome, but neither was confirmed by the data ($t(339) = 1.04$, p = 0.30), see Figure 3.6 for the overall regression of all participants' answers on "Luck" on "Importance".

All three groupings had relatively low correlation with the perception of luck as a driver of outcomes (ρ_{low} = 0.19, ρ_{med} = -0.17, ρ_{hi} = -0.02), but one could interpret

Figure 3.6: Regression of Luck on Importance



Note: The regression in the Figure above of participants' answers to questionnaire item on the factor luck in games' outcomes versus the importance of the subject matter to participants. The data do not show any discernible relationship between the self-professed knowledge of participants and their estimations of whether luck plays a prominent role in games' outcomes.

the data as indicating that more participants with little knowledge estimate luck to be a factor in determining games. Those with a high self-professed knowledge see no correlation between luck and games' outcomes. Those with moderate knowledge seem to be those whose negative correlation could indicate that they think that the outcome of games can be predicted. Those who know more about the subject seem to realize that forecasting outcomes are more difficult! Therefore, one might assume that these participants of moderate knowledge could also be the most overconfident.

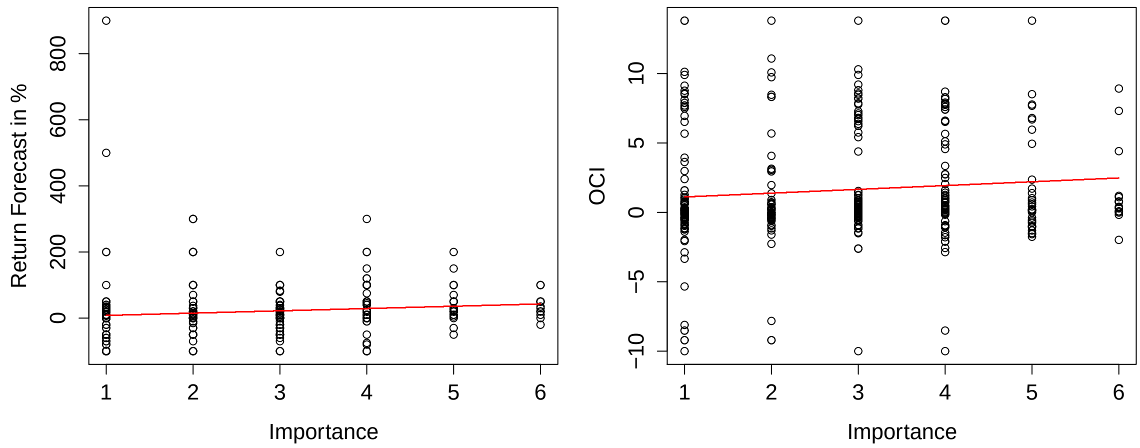
As described in the Introduction, one common form of overconfidence is observed when people believe they are better than the others at a task - the better than average effect (described for drivers in Svenson, 1981). Presson and Benassi (1996) find that people often consider themselves to have inaccurately high levels of prediction ability, and Langer (1975) finds that people think that they are able to control situations whose outcomes are, if fact, determined by chance and not by ability. On the basis of these results, one might expect participants in our market to overestimate their own abilities if they consider themselves knowledgeable about the subject matter or if they consider themselves experts because subject is important to them.

- **Hypothesis 5:** Participants for whom the subject matter is important or have high self-professed knowledge will also show higher overconfidence.
- **Result:** Interestingly, this hypothesis could not be confirmed beyond a tenuous relationship ($t(339) = 1.78$, p $= 0.08$, significant at a p $< .10$), see right-hand graph in Figure 3.7). Overconfidence in trading does not seem to be directly

linked to how important the subject is for participants. Also, the data cannot support any correlation of knowledge with overconfidence as discussed in the previous hypothesis. The three groupings of participants according to subjective expertise show hardly any correlation with overconfidence (ρ_{low} = 0.01, ρ_{med} = 0.004, ρ_{hi} = -0.08). However, there is a relationship between participants' subjective estimations of the importance of the subject matter to them and the level of their return estimates ($t(339) = 2.29$, p = 0.02, significant at the $p < .05$ level, see right-hand graph in Figure 3.7).

These results can be seen as a corroboration of hypothesis 3 - if participants do not select their expectations of trading results based on their subjective expertise, then their overconfidence ex post seems also not to be directly linked to subjective expertise.

Figure 3.7: Regressions of Importance on Performance Forecasts / OCI



Note: The regression in the left Figure shows that there is no discernible relationship between participants' forecasts of their expected performance (return) and the importance of the subject matter. The right-hand Figure shows that overconfidence as measured by the overconfidence index OCI is not predicted by the relative importance of the subject matter to participants.

The graphs in Figure 3.7 show the correlations between the importance of soccer for participants and with the participants' forecasts of performance (left graph), and with their ex-post overconfidence (right graph).

Who, then, are the overconfident traders, if not those with little knowledge, those with lots of knowledge or those with an interest in the subject matter? It turns out that the group of participants who answered that they have high confidence in their ability to predict the outcome of games are in actual fact the most overconfident group of participants as viewed ex-post by comparing their predictions with their actual

Table 3.9: Subjective Expertise in Prediction

Expertise	No.	ρ(Luck)	ρ(OCI)
Low Expertise	119	2.43%	−7.07%
Medium Expertise	168	5.16%	0.97%
High Expertise	53	13.61%	60.65%
Overall Correlation	340	6.09%	3.35%

Note: The columns in the Table above contain three subsets of participants grouped by their professed level of soccer expertise, the number of participants in each grouping, the correlation between the average answers in that grouping and their answers to the questionnaire item Luck, and the correlations between the groups and their average overconfidence as measured by the overconfidence index OCI. The most interesting result is the high correlation between those participants who consider themselves very knowledgeable on the subject matter, with their subsequent propensity to overconfidence.

achievements via the oci. This can be recognized from the correlation analysis presented in Table 3.9.

Table 3.9 shows the relationships of different groups of participants' answers to their estimation of their own subjective expertise in prediction (low, medium or high) with the categories Luck (how much does luck have to do with the outcome of soccer games) and OCI (their ex-post index of overconfidence). The results show that the by far highest correlation is of participants who professed high expertise in predicting the outcome of games ex-ante also showed the highest ex-post level of overconfidence as measured by the oci comparing return estimates with achieved results.

Those participants who are in the high group of self-professed experts in predicting the outcomes of soccer games are proven to be the most overconfident. On the one hand, they have the highest estimate that luck plays a part in the outcome of games, but they exhibit by far the highest ex-post correlation with the overconfidence index.

This seems also to be corroborated by the fact that the group of participants with a low perception of their expertise in prediction exhibit a negative correlation with the overconfidence index.

So in conclusion, it seems to be that overconfidence depends not on whether one is male or female, not on how much one trades, not on one's age (in our case), and not on how much expertise one has on the subject at hand.

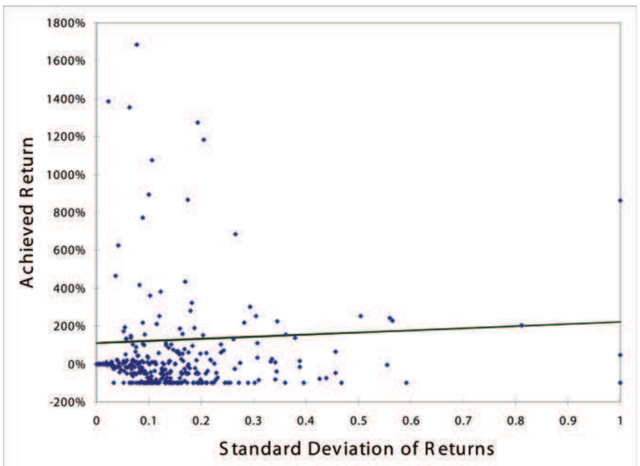
Rather, it seems that this is a first indication that overconfidence seems to be a character trait that is present from the beginning of trading throughout the entire market duration, rather than an acquired propensity, as suggested by authors such as Odean (1999). Increasing levels of (self-professed) expertise or even interest in the subject do not seem to predict that participants will show increasing levels of overconfidence. Overconfidence in our data is exhibited ex-ante by answers on subject-independent questions regarding personal expertise in prediction, and confirmed ex-post by the oci, and leads us to conclude that overconfidence is dependent on other variables than trading successes, diversification or variability of returns.

In order to investigate the reasons for overconfidence, we need to investigate the relationships within different groupings of over- or underconfident participants more. In the next chapter we introduce another measure to help us and perform some detailed analysis of participants' portfolios who are either very over- or very underconfident.

3.3.7 Identification of Trading Strategies

In order to investigate the reasons for overconfidence, we need to probe the relationships within different groupings over over- or underconfident participants more. In order to achieve this, we introduce additional measure for analysis and perform some detailed analyses of participants' portfolios who are either very over- or very underconfident. Can we identify a relationship between participants' risk-taking behavior and their overconfidence? In order to answer this question, we look at the individual portfolio risk, return and diversification.

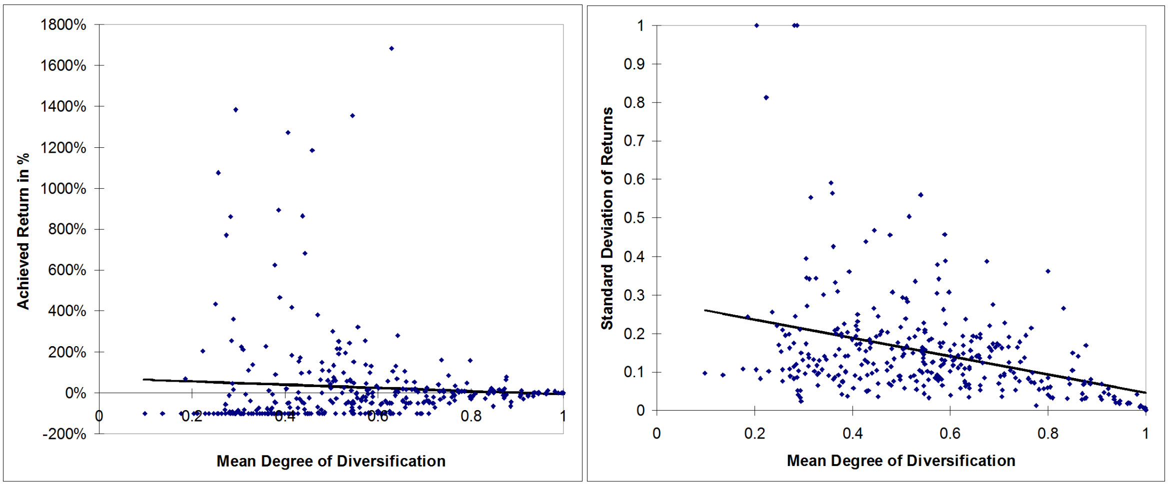
Figure 3.8: Risk - Return Relationships



As diversification is primarily used as a risk-reducing tool (see Markowitz, 1952), one first analysis (shown in Figure 3.8) plots overall return against risk measured as the standard deviation of returns in usual Markowitz fashion. In order to ultimately identify whether sources of overconfidence are manifested via trading strategies, Figure 3.8 shows two linked components of motivations for successful trading: risk and diversification/concentration.

The regression analysis in the left chart of Figure 3.8 shows that any relationship between risk and return of our participants is tenuous at best ($\rho = 0.07$, $t(339) = 1.21$, p = 0.23). Since there is no easily detectable relationship between risk and return, we analyze whether diversification had any impact on returns or risk, shown in Figure 3.9.

Figure 3.9: Risk - Return vs Diversification



As shown in the left-hand graph in Figure 3.9, there does not seem to be any major relationship between the degree of diversification and participants' returns ($\rho = 0.07$, $t(339)$ = -1.31, p = 0.19). However, the relationship between the DoD and the risk taken as measured by the standard deviation of returns shows a significant connection, as would be expected under a Markowitz regime ($\rho = 0.36$, $t(339)$ = -7.17, p $<$ 0.001). Even though diversification reduces risk, it seems that overall diversification is not a reliable indicator of return within our sample. Since there are too many outliers to derive a general conclusion from this data, so we need to identify different groupings within the broad mass.

Separate analysis of the decile with the most diversified and the least diversified traders is necessary to find potential trading strategies that might lead to overconfident behavior. Analysis of the most diversified and the least diversified traders shows significantly more variability in the returns of those who diversified less. Diversification within the portfolio did not guarantee returns (cash holdings did not earn interest and it was a zero-sum game), as evidenced by the mean returns in Table 3.10.

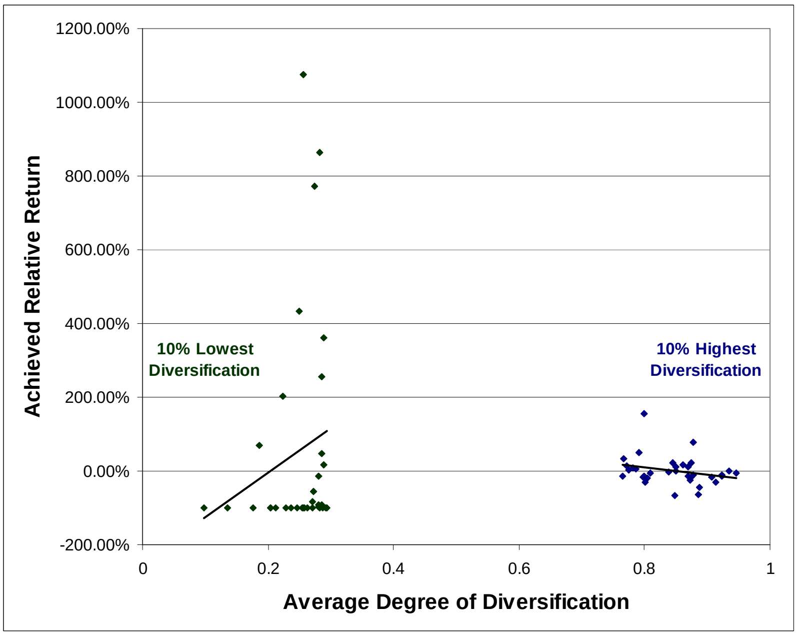
There is a large difference between the mean and the median returns in both groupings, but particularly within the non-diversifiers, which is explained by the large return outliers by the few (winning) participants. The non-diversifiers achieved a median return of almost -100%, indicating that most of the non-diversifiers lost everything and

Table 3.10: Portfolio Diversification Statistics

Diversification	Top 10% Diversified	Bottom 10% Diversified
Mean Return	0.27%	54.87%
Median Return	−4.74%	−99.60%
Minimum Value	−65.0%	−100.0%
Maximum Value	156.0%	1076.0%
Median Stdev	5.75%	12.0%
Median OCI	1.13	52.82
Mean Prediction Ability	3	4

a few made huge returns (see also on the left side of Figure 3.10, this shows the distribution of returns per degree of diversification). This is evidenced by the minimum and maximum values in Table 3.10, and the fact that the non-diversifiers were more than twice as risky (as defined by the standard deviation of returns) as the diversifiers.

Figure 3.10: Diversification - Return Relationships



Modern portfolio theory as initiated by Markowitz (1952) formalizes the risk-return tradeoff. A portfolio can only offer more expected return if the investor accepts more risk. Conversely, the more risk-averse the investor, the lower the expected return must be - there is no free lunch available. The driving factors under modern portfolio theory are thus asset allocation - which stocks to hold in the portfolio and in which proportion - and market timing - when to buy or sell these stocks (for probably the definitive overview of classical portfolio theory, see Litterman, 2003).

If trading strategies from a Markowitz (1952) viewpoint can be linked to risk/return, diversification/concentration and market timing, then the first point can be addressed insofar that we could not in our sample identify trading strategies based on risk - return correlation. Trading strategies based on decisions of whether to diversify or concentrate holdings seem to be confirmed by the analysis shown in Figure 3.10. The third point, market timing, will be addressed within its link to overconfidence in the subsequent chapters.

3.3.8 Connections between Overconfidence and Trading Strategies

Viewed over the lifetime of trading, it seems that both the most and the least diversified were overconfident on average, but that the least diversified group of traders was on average much more overconfident than the more diversified (the median OCI values are 1.13 and 52.8 respectively).

Diversification of individual portfolios or focusing on a few stocks can be viewed as types of investment strategies in addition to providing an indication of individual risk-seeking or -avoiding behavior. From the above results, the non-diversifiers turned out to be much more overconfident than the diversifiers. As these are overall ex-post results the interesting question behind this is the breakdown over time of this behavior: Did these results follow from trading strategies that were adhered to throughout the competition or can one observe learning behavior in the one or the other direction? For example, the following strategies could be possible: a) Follow through with a diversified strategy (no change in behavior), b) Follow through with an undiversified strategy (no change in behavior), c) Start with a diversified portfolio but become overconfident over time and reduce diversification (change in behavior), or d) Start with an undiversified portfolio but become cautious over time and increase diversification (change in behavior).

In order to identify whether ongoing trading results led participants to change their strategies (either learn from their losses or become overconfident through profits) first we identify the characteristics of the overconfident versus the underconfident participants. Then we analyze the returns, riskiness, degree of diversification and overconfidence of participants over time. This should provide an indication of whether over- or underconfidence was a property that was inherent in the trading strategies from the beginning or whether strategies were changed over time due to increasing over- or underconfidence. To achieve this we analyze the top and bottom deciles representing the most and the least ex-post overconfident traders (n = 34 in each group).

3.3.9 Does Overconfidence Change over Time?

In order to discover the reasons behind activity that could lead to overconfidence, we analyze the following parameters for both groups (top 10% most overconfident and bottom 10% least overconfident participants); see Table 3.11. By observing the following measures, which are re-computed for each participant after every single trade is made, we can provide an ongoing observation of risk, return, overconfidence or degree of diversification, rather than just at pre-determined intervals. The following measures are mainly used to establish our findings below:

1. Activity (act) - the number of times the participant actually traded.
2. Relative Cumulative Return (rel_cum_ret) - a participant's realized return in percent.
3. Mean OCI versus median OCI (mean_oci, median_oci) - the difference between the two measures shows the presence or absence of large outliers. To minimize the effect of outliers, the median OCI is used.
4. Number of times profitable versus loss-making trades were entered into per participant.
5. Relationship of number of times a trader has a rising degree of diversification divided by the number of times a sinking degree of diversification is observed (dod_ud, or no_dod_up / do_dod_down). This shows how often traders changed their strategies for diversification.
6. Relationship of number of times the oci rises divided by the number of times the oci falls (oci_u/d, or no_oci>1 / no_oci<1). This shows how often traders put their overconfidence into effect in the market by either buying or selling stocks. Many of the overconfident traders show a relationship of $+\infty$. The number of times their overconfidence index is > 1 (overconfident) is a positive number, and the number of times they were underconfident is zero.

(a) Comparison of the Most and Least Overconfident Participants

The data in Table 3.11 show the direct comparison of the discussed measures between the most overconfident and the most underconfident deciles. the left column lists the variables that were used for evaluation, the other columns show the averages, minimum and maximum values[11].

Table 3.11: Comparison of Overconfident versus Underconfident Decile

	Average		Min / Number < 0		Max / Number > 0	
Category	OC	UC	OC	UC	OC	UC
act	137.41	335.47	16	2	498	2017
rel_cum_ret	0.00	5.27	0.00	0.00	0.00	17.84
stdev	0.17	0.24	0.08	0.02	0.59	3.36
oci	185562.58	0.09	2666.67	0	99999.00	0.24
mean_oci	2318.28	0.40	11.00	0.00	22224.35	2.51
median_oci	1.72	0.38	0.72	0	3.94	2.59
profit_trade_no	56.38	181.47	4	1	219	1371
loss_trade_no	84.74	160.18	14	9	299	1075
tr_u/d	0.81	1.46	6	0.05	28	8.44
no_oci>1	98.82	42.53	12	0	422	389
no_oci<1	42.29	299.41	0	1	424	2023
oci_u/d	5.99	1.48	6	0	28	19

Note: OC: overconfident, UC: underconfident
act: Activity, the number of times the participant actually traded
rel_cum_ret: Relative Cumulative Return, a participant's realized return in percent
mean_oci, median_oci: Mean OCI versus median OCI - the difference between the two measures shows the presence or absence of large outliers. To minimize the effect of outliers, the median OCI is used
profit_trade_no, loss_trade_no: Number of times profitable versus loss-making trades were entered into per participant
dod_ud, or no_dod_up / do_dod_down: Relationship of number of times a trader has a rising degree of diversification divided by the number of times a sinking degree of diversification is observed. This shows how often traders changed their strategies for diversification
oci_u/d, or no_oci>1 / no_oci<1: Relationship of number of times the oci rises divided by the number of times the oci falls. This shows how often traders put their overconfidence into effect in the market by either buying or selling stocks. Many of the overconfident traders show a relationship of $+\infty$. The number of times their overconfidence index is > 1 (overconfident) is a positive number, and the number of times they were underconfident is zero

Table 3.11 shows that the number of trades measured by act (activity) is higher on average (double the number of trades) for the underconfident decile, meaning that overconfidence did not imply increased trading activity (significant at 5%: $t(33) = -2.3$, p = 0.02, two-tailed).

The average cumulative return is quite different between the two groups. The

[11] The data showing trade-by-trade development of all indicated measures and calculations is available from the corresponding author on request

overconfident decile has an average cumulative return of almost zero, whereas the most underconfident group scores highly on this count at an average multiple of 5.27 (significant at 1%: $t(33)$ = -5.84, p < .001, two-tailed). The standard deviation of the underconfident group is not significantly different from that of the overconfident decile ($t(33)$ = -0.67, p = 0.50) - it seems underconfidence can be exhibited at the same risk level.

One major difference between the over- and the underconfident group can be seen in the median oci. The overconfident group is on average overconfident for the long-term (the average of all their long-term overconfidence is 1.72, indicating long-term average overconfidence), and the same parameter for the underconfident group is 0.38 (significant at 1%, $t(33)$ = 11.99, p p < .001). Of the 34 overconfident traders in the decile, only 5 have a median oci of under 0.97. Of the underconfident group, only 4 have an oci that is greater than this value, which indicates that over 85% of both the most overconfident and the most underconfident participants exhibit long-term over- or underconfidence. Analysis of the median oci relative to the mean oci provides an insight into whether the average is dominated by outliers or is consistently high or low. In our case this means that even though the median oci for the top decile is quite different from the corresponding average oci, it still shows persistent overconfidence for the top decile (as it is > 1) and persistent underconfidence (> 1) for the bottom decile. So here again, overconfidence was not learned, but exhibited (on average) from the beginning, as was first mentioned in the previous chapters analyzing the individual deciles.

To confirm this, we look at how often the participants in these groups invested in profitable trades versus loss-making trades and what the ratio between the two values is. We would expect our persistently overconfident traders to show more loss-making trades than profitable ones, which is evidenced in Table 3.11 (an average of 56.38 profitable and 84.74 loss-making trades, yielding a ratio of 0.81). The underconfident traders show the opposite: an average of 181.47 profitable and 160.18 loss-making trades for a ratio of 1.46.

(b) Analysis of the Top Decile of Overconfident Traders

Does overconfidence manifest itself in trading behavior from the beginning or does it lead to participants changing their strategies as a result? The analysis of the top 10% overconfident traders shows that different strategies were obviously pursued. As the participants in this bracket are all very overconfident versus their expectations, i.e.

their realized returns fall well short of their expected returns, where did their strategies go wrong?

Results from analyzing the returns, risk, diversification and oci are unexpected, based on results published in previous literature, such as Benos (1998), Barber and Odean (2001), Daniel et al. (2001), or Scheinkman and Xiong (2003). We find hardly any indication of systematic learning behavior, as authors such as these document. Rather, participants mainly “stuck to their guns”.

Based on the development of overconfidence over time, we look at the percentage of time that a participant (from the top decile) was overconfident by identifying per participant how often their OCI rose or fell measured after each trade. From this we can tell whether a participant was predominantly over- or underconfident.

Based on how much of the trading horizon a participant was overconfident, we can identify different types of trading or investment behavior. We separate the top decile of overconfident participants into three distinct types based on the predominance over time of overconfidence and additionally, when the overconfidence was present during trading, as well as analyzing whether the portfolios were kept constant over time or were churned (rolled over frequently). The three types of overconfident behavior we define as overconfident traders, overconfident investors, and last-minute speculators.

For example, if a participant was constantly overconfident and kept his portfolio relatively unchanged for the duration of the tournament, we would classify this as overconfident investment behavior. For some reason this investor has determined that the stocks in his portfolio are superior to others, despite information to the contrary.

In contrast to this, if a participant was overconfident for most of the time, but turned over her portfolio constantly, she would be classified as an overconfident trader (as opposed to an investor). She would be of the (incorrect) opinion that she was able to pick short-term winners better than the rest of the market.

The third classification of overconfident behavior results from less constant overconfidence than the other two categories, a major change to portfolio asset allocation at the very end of the trading horizon, and a large loss with resulting overconfidence peaking after the last match was conducted - a last minute speculator (who lost). From these three types of overconfident behavior we classify all participants within the top decile.

Overconfident Traders:

Their trading behavior exhibits consistent overconfidence over most of the investment horizon. They realize their losses immediately. We identify such traders by observing their oci over time. If the percentage of (number of times oci rises) / (number of times oci falls) shows that the number of times the trader is overconfident is $> 75\%$, then we conclude that this trader was predominantly overconfident over the investment period. We identify 23 such candidates out of the 34 in the decile. A possible explanation for this might be that overconfidence for these traders (68% of the 34 top overconfident participants) is a trait that is not acquired during trading but rather is present from the very beginning. For for most of these participants (18) - the oci did not once leave the overconfidence region, i.e. overconfidence was present from the start. Overconfidence bias predominant.

Overconfident Investors:

They are either not or only slightly overconfident for most of the trading horizon as defined by the result of their realized profits and losses. At the end they lose out from unrealized paper losses they have pulled along for a long time that are realized at the end. These are investors who hold on to their losing stocks for too long. We define such investors as those who show a large increase in oci at the very end (when the contracts expire), but have built their portfolios over a longer period of time, i.e. not at the last minute. We identify 10 such investors (out of 34 in the decile). The mean oci excluding each participant's final redemption by the exchange is 1.42. The mean oci of each participant's portfolio including final redemptions is 207,841. This documents the huge jump in overconfidence resulting from positions held within the existing portfolio (not entered into at the end). We hypothesize that these participants are actually overconfident in the same manner as the overconfident traders from the previous point, but only realize their overconfidence at the end. Loss avoidance bias predominant.

Last-Minute Speculators:

They have traded more or less cautiously for much of the time. On the last day, they throw caution to the wind and pick one of the two remaining teams to invest all their spare cash into. Obviously, they picked the wrong team! As indicated before, activity (number of trades performed) is not a good indicator of overconfidence in our study. The group of most overconfident traders is actually on average less active than the most underconfident group (average of 137 versus 335 trades per trader). We identify these traders by observing whether a last-minute trade caused a huge jump in oci (in contrast to the overconfident investor, where the last-minute jump in oci is not caused by a last-minute trade, but by a buildup of a position over time). The entire portfolio has been sold off and focused on one remaining stock (which then represented the losing team). In the overconfident decile, we identify only one such last-minute speculator (out of 34 in the decile). This participant's oci jumped from a mean of 1.34 before entering into the final trade to an oci of 3333 after the final trade. One could argue that this group contains those who are not really overconfident in the narrower sense of the word, but rather reckless in their risk-taking behavior.

Concluding the analysis of the most overconfident decile of participants, we observe that 33 of the top 34 overconfident traders seem to be overconfident for a large portion of the investment horizon, whether their overconfidence is immediately realized in sub-par returns, or whether they accrue paper losses for a long time and only realize them at the end. Among the 34 top overconfident traders, only 1 was a last-minute risk-taker, the others were much more systematic concerning their investment behavior (and thus also their implicit or explicit overconfidence). These findings from the group of most overconfident participants seem to lead to the conclusion that maybe overconfidence does not build up over time, but is rather an inherent behavioral characteristic.

(c) Analysis of the Bottom Decile of Overconfident Traders

The analysis of the bottom decile of participants ranked by their overconfidence, a similar but not identical picture emerges. We separate groupings by strategy analogously to the most overconfident decile.

Underconfident Traders:

Of the 34 traders in the bottom decile representing the least overconfident traders, a full 23 of 34 (68%) never left the region of underconfidence, calculated by the number of times oci > 1 divided by the number of times oci < 1, i.e. the traders were never overconfident.

Analogously to the "overconfident traders" from above, one can also identify mostly underconfident traders (n = 6) as those whose number of occurrences of (oci $<$ 1) divided by the number of occurrences of (oci $>$ 1) is $>$ 75%, indicating persistent underconfidence. As 29 of 34 participants qualify as at least mainly underconfident (85%), it seems that underconfidence is also a habit that is present from the start, rather than acquired during trading.

Underconfident Investors:

In the bottom decile representing the most underconfident traders, there were no underconfident investors to be observed in analogy to the overconfident investors from the most overconfident decile.

Last-Minute Speculators:

This group of underconfident traders is in analogy to the overconfident group who are unlucky with their last-minute trades, one could call these participants the lucky few - they are mainly overconfident (number of oci $<$ 1 / number of oci $>$ 1 is $<$ 75%), but were lucky enough to bring their overconfidence level down by happening to be lucky at a large trade. We identify 5 of these participants (n = 5 of 34).

So in analogy to the analysis of the top decile of overconfidence, we can identify several strategies within the group of participants exhibiting underconfidence: those who are always underconfident, those who are mostly underconfident and the lucky few. Interestingly, the number of traders in the first two groups (29 of 34, or 85%) do not waver from their habit of underconfidence, in similar fashion to the traders who are mostly overconfident as a rule. Maybe underconfidence is also previously established, rather than acquired as the opposite facet of confidence as a characteristic habit?

3.4 Conclusion

Basing an investment market on sports trading has the advantage that the outcomes of games are easily understood and that the dynamics of trading or investing are of interest to trading on the capital markets. Smith, quoted in Dowie (1976) states that *"... institutional forms of betting are of scientific interest for two reasons. They yield behavioral implications for individual decision-making under uncertainty. Furthermore,* *since these betting schemes give rise to wager markets with equilibrating functions similar to ordinary security (or commodity) markets, they afford opportunities for the study of market mechanisms under a widened class of contingency and institutional conditions".* Similarly to forecasting on the capital markets, research has shown that forecasting ability in sports settings is also prone to error (Andersson et al., 2005). Overconfidence can be a prime source of this failure to predict outcomes reasonably correctly.

Previous research such as by Gervais and Odean (2001) or Odean (1999) indicates that overconfidence in a trading or investment setting can arise as feedback from trading successes, leading to an over-estimation of one's own abilities and increased trading turnover. Our investigation into this phenomenon led us to analyze various differing aspects of potential overconfidence, for which we also develop two new measures, a Degree of Diversification (DoD) and an Overconfidence Index (OCI). The initial investigation was able to show that overconfidence was indeed widespread amongst the participants of our study that was based on a prediction market trading on sports teams.

Participants who actively traded via the on-line platform had considerably diverging expectations versus results. More investors achieved below-average results than they expected - 22% expected to lose money, whereas over 62% actually did. Conversely, 78% expected not to lose money, but only 38% were able to finish above zero return. On an absolute level, this confirms expectations that, particularly within a sports betting or trading system, overconfidence can easily be observed.

Combining this information with the correlation of ex-ante responses to a questionnaire shows that those who were initially (before the study began) very confident of their abilities continued on to exhibit overconfident behavior, greatly underperforming versus their expectations. Conversely, those who were initially cautious regarding their prognosticative abilities turned out to have a negative correlation with measured overconfidence.

Portfolio theory as initiated by Markowitz (1952) and psychology see risk-taking behavior from different perspectives. Classic portfolio theory views reduced or non-existent portfolio diversification as risky behavior, as increased diversification lowers portfolio risk (measured as the standard deviation of returns). Our data does not show a correlation between this behavior and the psychological phenomenon of overconfidence, which psychological literature (such as e.g. Grinblatt and Keloharju, 2009) relates to sensation-seeking behavior.

Our research into the trading behavior of those participants who ended up being the most or the least overconfident yielded rather surprising results; it turns out that in our study participants did not learn overconfident behavior as a result of trading successes. Neither did they learn to switch to risk-averse portfolio construction by diversifying more after trading losses. Rather, when attempting to discern trading strategies within the most overconfident and the most underconfident participants, most did not change strategy much. We hypothesize that most participants clung to pre-established opinions and conducted their trading strategies correspondingly. Last-minute speculators who abandoned a previous trading strategy and threw everything into one last-gasp hope were not observed very often when analyzing the most or least overconfident groupings.

We conclude from these results that overconfidence was (at least in our experiment) inherently present from the beginning and was not systematically reinforced or generated by positive or negative feedback strategies, successes or failures. Overconfidence was exhibited by many (over 60%) of the participants, but did not result in measurable different trading strategies, stock turnover (churning), risk-taking behavior or diversification. Underconfidence was just as stable and led to ca. 85% of the least overconfident participants to exhibiting roughly the same behavior over the course of the study; "underperformance" of the "overconfident" is just as stable as "overperformance" of the "underconfident". From this one might conclude that overconfidence (at least concerning trading) is a function or "nature" rather than of "nurture", i.e. constitutes a character trait rather than arising from feedback concerning trading successes or the lack thereof.

Why then are so many participants so overconfident of their ability to achieve above-average returns? It almost certainly has nothing to do with the discussion regarding for-real-money versus for-play-money prediction markets. According to other authors, traders or investors are not necessarily only motivated by real money in order for them to trade or invest to the best of their abilities (see e.g. Servan-Schreiber et al., 2004). Wolfers and Zitzewitz (2004) describe how prediction markets, such as the one we implemented, can work efficiently if *"... a motivation to trade exists"* (Wolfers and Zitzewitz, 2004). The motivation to succeed was present as demonstrated by the Activity Index results which measured how many trades each participant made during the investment horizon. The participants who actively traded showed an average number of trades of 130, i.e. almost 11 per person and trading day; the participant with most trades had 2017, or almost 75 per day of participation. The makeup of the participants may have had an influence, as most participants were between 20 and 30, and either university students or employees of a consulting company.

What are the implications for academic research? Previous research has often focused on miscalibration as an indication of how well people form confidence intervals around their estimates (see e.g. Lichtenstein et al., 1982). Perhaps further research into whether expectations of successes on financial markets related to participants' inherent personal characteristics could shed light onto why confidence seems to manifest itself as stable over- or underconfidence in trading settings.

Further research would be interesting to analyze whether overconfidence really is related to personality characteristics rather than being learned "on the job". In particular, as we were unable to confirm changes to over- or underconfident behavior during trading - the top and bottom decile of over- and underconfident participants were remarkably stable; the amount of trading activity was also unrelated to overconfidence in our experiment. From this evidence it suggests itself to perform further research within personality theory as a field of psychology. Within personality theory, trait theory describes characteristics of personality that shape characteristic patterns of behavior (see Eysenck, 1990). Traits are defined as *"... enduring qualities or attributes that predispose individuals to behave consistently across situations"* (Gerrig and Zimbardo, 2010). There is a discussion in psychology literature as to whether traits act as predispositions that cause behavior, or whether they are only valid as descriptive patterns of observed behavior (see e.g. Allport and Odbert, 1936; Allport, 1937, 1966, on the one hand, and Eysenck, 1990; McCrae and Costa, 1997, on the other hand). Behavioral consistency across situations can be found by identifying the psychologically relevant features of situations (see e.g. Mischel, 2004). From this vantage point, further reasearch could be addressed at identifying whether overconfidence (e.g. in finance) can be found in consistent form across situations.

As many real-life players on the financial markets are actually rewarded for their overconfidence (or at least their outward portrayal of overconfidence) it is increasingly important to recognize this behavior - as most participants of our study found out to their detriment.

Chapter 4

Forecasting Overconfidence

4.1 Introduction

Studies of behavioral biases now have a richly documented history since Daniel Kahneman and Amos Tversky first applied psychological methods and insights to economic questions from the early 1970s onwards. Many behavioral biases have been identified and tested in economic and financial settings, overconfidence being one of the most prominent observable biases. Overconfidence can be responsible for inefficient, sub-par or less-than-expected results, depending on the focus of the research in question. Plous (1993, p. 217), states of overconfidence that *"[n]o problem in judgment and decision making is more prevalent and more potentially catastrophic"*.

In psychological research, overconfidence reflects a systematic discrepancy between expectations and reality and leads to subjects feeling worse about negative outcomes than their well-calibrated counterparts (McGraw et al., 2004). Overconfidence among managers leads to poor long-term performance (see e.g. Doukas and Petmezas, 2007) or incorrect corporate finance decisions (see e.g. Ben-David et al., 2007). In an investment or trading setting, such as documented in Odean (1998b) or Barber and Odean (2000), overconfident market participants experience diminished returns amongst other phenomena.

If overconfidence leads to unwanted results, one would expect there to be research into how to avoid overconfident behavior or at least into how to recognize it before the fact. In psychological academic research, the issue of identifying overconfidence and applying methods such as debiasing to subsequent subjects has shown that the effects can be mitigated under certain circumstances (see e.g. McGraw et al., 2004), but Fischhoff (1982) reports on the difficulties associated with debiasing as a reliable technique.

Research by Hilton (2001) gives pointers on how potentially to deal with biases in trading and investment, but on a very general level. Dealing with behavioral biases has surfaced in mainstream popular publications, on a general level e.g. by Ariely (2008) or with finance in particular in mind, such as Nofsinger (2011) or Belsky and Gilovich (1999). The latter authors give the following advice to counter effects of overconfidence bias (Belsky and Gilovich, 1999):

1. Investor, know thyself: The recommendation is to educate oneself on the topic, and to backtest one's (potentially overconfident) decisions by writing them all down and evaluating a year later which percentage of the investment decisions would have performed well.
2. Take 25% off the top and add 25% to the bottom: The recommendation is to add an "overconfidence discount" as an add-on to the values of decisions.
3. Get a second opinion: The recommendation is to solicit opinions, not on the decisions themselves, as the other person is presumably also overconfident, but on the decision process.

Within a trading or investment context, the unwanted results due to overconfident behavior cost either individuals or institutions large amounts of money. Excessive trading due to overconfidence results in excessive transaction costs (Odean, 1999; Barber and Odean, 2000). Diminished returns due to overconfident trading or investing behavior can cost individuals or institutions a lot of money in opportunity costs, see Gort (2009) for research on Swiss pension funds. One could argue that when it comes to avoiding the costs incurred by behavioral biases to individual investors, it is up to them to recognize and correct their behavior, maybe by following the recommendations in popular books such the ones mentioned above as examples.

However, in an institutional setting, avoiding realized or unrealized losses is a very important task intrinsic to the institution. Effective risk management whose purpose is to identify and to limit losses is essential and is therefore also mandated by supervisory authorities. In the United States financial institutions are subject to many federal and state laws, and in Europe to the current Basel regulations[12] (see www.bis.org) or the corresponding European Union (EU) directives (see http://ec.europa.eu), all of which mandate extensive risk management responsibilities. Shareholders and boards also have an inherent commercial interest in institutions making the most of their risk management capabilities. Financial Institutions are also required to comply with internal audit findings and recommendations concerning the ability of their risk management personnel and procedures in shoring up potential loss-making threats.

Traditional risk management focuses on many areas such as credit, market, operational, or legal risk which are detailed by authors such as Crouhy et al. (2000). The measurement of changes in returns due to fluctuations of underlying market variables such as stock prices, foreign exchange rates or interest rates is traditionally assigned to the area of market risk (see e.g. Jorion, 2006). Even though behavioral finance has been researched in relation to individual and institutional trading and investment as outlined above, to the best of the author's knowledge, no such research exists in respect to risk management.

The existing literature on overconfidence applied to finance has focused primarily on detecting the presence of overconfidence in one form or the other. Detection of overconfidence is a first necessary step, but the question remains whether this information, as well as potential techniques, can be identified not only to detect overconfidence, but to detect it early enough to do something about it. The existing literature on overconfidence neither addresses this aspect within an investment nor within a risk management context.

This paper aims to fill this gap in the literature and empirically evaluates the possibility of forecasting overconfidence with risk management applications in mind. The empirical analysis was performed using experimental markets that have been shown to represent the relevant underlying points from real-life markets (viz. earlier chapters).

In the following sections, we outline how we identified methods to forecast overconfidence using multivariate statistical techniques such as multiple regression, factor

[12]The Basel Committee on Banking Supervision within the Bank for International Settlements produces recommendations on financial regulation and capital requirements, which are subsequently transformed into law by countries' and regions' legislative authorities. The rules integrate capital standards with national regulations by setting the minimum capital requirements of financial institutions with the goal of ensuring financial institutions' liquidity. Basel II is the second of the Basel Committee on Bank Supervision's recommendations, and unlike the first accord, Basel I, where the main focus was on credit risk, Basel II focuses on how much capital financial institutions must put aside. Banks are required to furnish enough capital to reduce the risks associated with its investing and lending practices.

analysis and correspondence analysis. In the Discussion Section, we describe where we find starting points for pinpointing how to forecast those individuals who might be more prone to overconfidence than others. Based on this methodology, we indicate how risk management departments might adapt their procedures to base early-warning systems on overconfidence forecasting in order to catch potential losses due to overconfidence bias before they mount up.

4.2 Method

4.2.1 Research Objectives

In previous research on behavioral finance in general such as in e.g. Kahneman et al. (1991), Thaler (1987b) or Thaler (1987a) and overconfidence in particular (see e.g. Gervais and Odean, 2001, Barber and Odean, 2001, or more recently, Gort, 2009), behavioral biases have usually been tested either by evaluating data using classical portfolio parameters such as expected utility, expected returns, standard deviations of returns, portfolio performance or distributions of values; or in a completely different fashion by comparing declared intentions via questionnaires with ex-post data, such as in e.g. Grimes (2002) or Glaser et al. (2007). Classical portfolio parameters have the advantage that they are well-known and comparable. Obtaining professed intentions via questionnaires has the advantage that questionnaires are readily customizable and are one of the main methods for comparing ex-ante indications with ex-post observed behavior. The main disadvantage of calculated portfolio values lies in the fact that they are only observable after the fact, and are not available ex-ante to determine people's intentions. Questionnaires, on the other hand, are a useful tool for evaluating test subjects' prior expectations and intentions. In order to test validity (i.e. measuring what was intended to be measured), questionnaire items need to be carefully structured with the subsequent portfolio activities in mind.

As the comparison of returns, volatility, activity and diversification seem not to be conclusive differentiators between over- and underconfident behavior (see Section 3 of Chapter 3: Overconfidence in Soccer), we focus on finding relationships between overconfidence and the answers to the questionnaire completed by each participant (for details of the questionnaire, see the Methods Section of Chapter 3: Overconfidence in Soccer). We use portfolio statistics such as log-returns and standard deviations of log-returns, as well as our own derived variables (degree of diversification and the log-overconfidence index) in combination with participants' answers to the questionnaire items in order to identify the driving factors behind overconfident behavior.

In order to find potential common characteristics, three methods were applied: regression analysis to determine potential relationships between log-overconfidence and the answers to the individual questions, exploratory factor analysis to find common characteristics, and correspondence analysis to render more precisely final confirmatory evidence of commonalities. The data from the correspondence analysis was then analyzed to permit the driving factors behind overconfidence to be identified.

With these data, we are able to identify those factors that contribute the most to subsequent overconfident behavior. By identifying those participants who scored the highest on questionnaires provided before trading commences, it was possible to identify those who would prove to be the most overconfident after the fact. Migrating this methodology from an experimental market to real-life trading situations should provide risk managers with tools to identify potential overconfident traders. By focusing on those traders a risk management department would be able to implement an early warning system and prevent potential losses caused by overconfident trading behavior before the fact.

4.2.2 Data

An on-line electronic trading platform was established which allowed study participants to trade on the outcomes of soccer games during the European Soccer Championship 2008. The platform admitted students at two universities in Vienna (37.35% of the final sample of participants), as well as employees (27.35%), clients (15.0%) of an international consulting firm and other professionals (20.30%).

The broad professional and demographic distribution provided a realistic environment for trading stocks as one would in real-life markets. Supply and demand determine prices, investors have to deliberate the consequences of posting bid or offer trading indications and the price discovery mechanism condenses the varied information into one figure, the current price.

Before each participant was permitted to trade, a short psychological questionnaire was requested to be filled out online, as well as demographic data such as age, sex, occupation and residence.

Of the 495 persons overall who entered the trading platform to place trades, 61 did not provide performance estimates or answer the psychological questionnaire. The 434 participants who provided the information required for study participation formed the initial study sample. Of all the participants in the study sample, 340 actively traded while 94 participants did not perform any trades, the 340 active participants formed the final study sample. Of this sample, most participants were male (63.24%) and between the ages of 20 and 29 (56.18%).

The original data consisted of a file containing 50,047 individual lines of transactions; from these transactions, incorrect or partially missing data were cleansed, cancelled orders were removed, and all transactions from one participant were removed who had proved to have cheated and violated the general terms and conditions of the market. From each order containing a buyer and a seller, the price of the transaction was extracted and used to form each of the 16 teams' price series over time; this resulted in a file containing 48,757 individual transactions. These transactions were assigned to each participant in his or her portfolio at each instance in time.

4.3 Measures

In order to find commonalities within the data that would enable forecasting of overconfidence, several exploratory methods were used. Overconfidence has been identified in the previous chapters by using 5 questionnaire answer data from the 340 participants, as well as portfolio measures such as returns, standard deviations of returns, an overconfidence index OCI, and a diversification measure DoD per participant. The previous chapters found that overconfidence as measured by the OCI could not adequately be represented by combinations of the analytic portfolio measures mentioned before.

However, we attempt to identify the common factors that correlate with overconfidence from all available parameters. We perform several different statistical methods in order to find either linear effects, commonalities underlying the data, or clusters in the data that are difficult to detect by other means. If reliable explanations for overconfidence can be found with one of the statistical methods of

- Linear effects of variables (by multiple regression analysis), or
- Common factors underlying describing data with a reduced variable set (factor analysis), or

- Clustering of data within large sets that have common characteristics (correspondence analysis),

then we can use this description of which variables determine overconfidence to attempt to forecast overconfident behavior. We perform all three of these methods in order to find at least one method that adequately describes the determinants of overconfidence in our data set.

First, we perform a multiple regression of the overconfidence index on the data for all participants in all 9 questionnaire answers and portfolio parameters (return, risk, diversification) in order to detect linear effects of the observable or calculated parameters on overconfidence.

Next, we perform a factor analysis on the data in the 9 previously mentioned variable categories plus the overconfidence index per participant. The factor analysis will show any common factors underlying the variables if there are any present in the data that would describe overconfidence by loadings on the variables.

Finally, we perform a correspondence analysis on the data within the 5 questionnaire answer categories and attempt to determine clustering in the data by questionnaire item. This will show clusters of answer data related to overconfidence effects, if they are present.

For the first multiple regression analysis and the factor analysis, the data consisted of the log-OCI as the dependent variable, and the questionnaire responses, as well as the forecasted returns, realized returns, degree of diversification and standard deviation of returns as explanatory variables. The questionnaire comprised 4 items which were measured on a 6-part Likert scale[13], see Likert (1931). We use a 6-point scale rather than the original 5-point scale in order to exclude the neutral point and force a decision. For arguments for and against inclusion of a neutral point in Likert scales, see Dumas (1999) or Eysenck (1998). The final sample number of participants was 340.

Detailed analysis in the correspondence analysis and the subsequent analysis of those results based on the results from the two methods above used the log-overconfidence index as the dependent variable and the questionnaire answers as independent data.

[13]The Likert scale requires respondents to a questionnaire to make a decision based on their level of agreement with a statement, generally on a five-point scale (ie. Strongly Agree, Agree, Disagree, Strongly Disagree)

4.3.1 Regression Analysis

The regression analysis is used to find potential linear interrelations between the base data and overconfidence. An initial exploratory multiple regression analysis was performed on all 9 relevant data categories in order to find out which response variables might be responsible for linear changes in log-overconfidence.

The log-overconfidence index was regressed on the questionnaire responses and the calculated return and risk data according to the following multiple regression equation:

$$logOCI_i = \alpha + \beta_j Q_{ji} + \beta_k P_{ki} + \epsilon_i \tag{4.1}$$

where for each participant i, each Q_j includes the four questionnaire items Importance, Knowledge, Prediction and Luck. The regressor P_k encompasses the portfolio values return estimate, realized return, mean diversification and standard deviation of returns.

We analyze the potential predictive quality of the ex-ante questionnaire items and the ex-post portfolio values together and independently of each other. Hypotheses for the relationship between overconfidence and the questionnaire items were already tested in Chapter 3 (Overconfidence in Soccer). The relationships between the portfolio values and resulting overconfidence were also analyzed in Chapter 3.

4.3.2 Exploratory Factor Analysis

In order to identify potential commonalities between the questionnaire responses and subsequent overconfidence, we perform a factor analysis. For this analysis we include all questionnaire and result data as in the full model regression analysis outlined in Equation 4.1, using factor analysis to reduce the 9 variables comprising the 5 questionnaire items.

The data comprises 340 sets of participant answers on a 6-part Likert scale comprising the following questions:

1. The importance of soccer (”compared to your other interests, what importance does soccer have for you?”) was assessed on a six-point Likert scale from 1 (”very little importance”) to 6 (”very high importance”),

2. Supposed subjective expertise on soccer ("compared to others, how well versed are you in soccer") was measured on a six-point Likert scale from "much worse than others" to "much better than others",

3. Supposed subjective expertise on predicting the outcome of soccer games ("compared to others, how well can you predict the outcome of soccer games" was measured on a six-point Likert scale from "much worse than others" to "much better than others", and

4. The supposed influence of luck on soccer game outcomes ("according to your opinion, which influence does luck have on the outcomes of the games of EURO 2008") was measured on a six-point Likert scale from "very little influence" to "very high influence".

5. The questionnaire asked participants to give a numeric estimate (in percent) of how their capital at the beginning would develop over the course of their participation. Participants were informed that, for example, "+50%" means that they win an additional 50% of their seed capital, "0%" means that their capital at the end will be as high as their seed capital, and that "-50%" means that they will lose 50% of their seed money. Instructions on the impact of trading on returns was posted on-line in the FAQ section of the platform.

Additionally, the results of portfolio calculations per participant were included:

6. Realized returns,

7. Mean diversification (the Degree of Diversification value, see Equation 3.2 in Chapter 3),

8. Standard deviation of returns (risk),

9. and the overconfidence index value per participant (see Equation 3.1 in Chapter 3).

The factors that were identified for this variable set are listed in Table 4.5 could be reduced to a smaller set containing potential "underlying" variables concatenated into common factors depending on their common correlations (for explanations of factor analysis, see e.g. Kim and Mueller, 1978b or Kim and Mueller, 1978a).

The varimax (orthogonal) rotation method is used (the varimax rotation method maximizes the variance of the squared loadings for each factor and is usually the preferred standard orthogonal factor rotation method and therefore the default option in statistical packages such as R, orthogonal rotation being the preferred objective factor rotation methodology rather than graphical rotation or oblique rotation); factor extraction was performed using maximum likelihood factor extraction[14].

How many factors to retain is not determined by an absolute rule, it depends on the goals of the analysis. There are several methodologies for determining the required number of factors. Cattell (1966) proposes the use of scree graphs to decide on how many factors or principal components to retain. The scree graph involves plotting the eigenvalues and finding a point where the line joining the points reduces steepness after a point k. Then, k components or factors are retained. Kaiser (1958) suggests dropping those components with eigenvalues that are less than one, as these contain less information than a single standardized variable whose variance is one. An additional methodology implemented in the statistics software R is the use of the Bayes Information Criterion (BIC) to determine the number of factors to retain (Revelle, 2011).

4.3.3 Exploratory Correspondence Analysis

Correspondence analysis provides a very useful graphical approach derived from scatter plots as spatial maps of interrelated data. It is a multivariate technique related to principal components analysis and factor analysis, in which the total variance is decomposed in order to arrive at a lower-dimensional representation of the variables that allow a reconstruction of most of the variance/covariance matrix of variables. As in principal components analysis, the idea in correspondence analysis is to reduce the dimensionality of a data matrix and visualize it in a subspace of low-dimensionality, commonly two- or three-dimensional data summarized in contingency tables or any other table of nonnegative ratio-scale data for which relative values are of primary interest (Nenadic and Greenacre, 2007).

All three methods have the common aim of reducing the dimensionality of a variable set, but there are differences between the techniques. Principal components analysis

[14]The objective of the maximum likelihood method in factor analysis is to find the factor solution which best fits the observed correlations. Other methods of factor extraction include minres (using the Ordinary Least Squares methodology to find the best fit) or the principal axis method, in which an eigenvalue decomposition of a correlation matrix is performed and then the correlation between the k-common factors for each variable are maximized. Authors such as Kim and Mueller (1978a) describe the maximum likelihood method as the most preferable methodology.

decomposes the total variance of the correlation matrix into the product of three matrices: the matrix of eigenvectors, its transpose, and the diagonal matrix of the associated eigenvalues. The correlation matrix can be decomposed into the principal component loading matrix with principal components as columns. It can then be approximated to any degree desired by retaining k principal components (Dunteman, 1989). Principal components analysis decomposes the correlation matrix without regard to an underlying model and does not distinguish between common variance and unique variance. Principal components can be expressed as linear functions of the variables or vice-versa.

Factor analysis, on the other hand, finds a decomposition of a reduced correlation matrix with a diagonal matrix containing the unique variances associated with the variables. Unique variances represent the part of each variable's variance that has nothing in common with the remaining p-1 variables. In factor analysis, since the variables are defined as a combination of common factors and unique factors, it does not provide a unique transformation from variables to factors, so different methods are possible. Factor analysis has an underlying model that rests on a number of assumptions, the key assumption being that the i-th variable can be expressed as a linear combination of hypothetical unobservable common factors plus a factor unique to that variable. In factor analysis, the initial factor solution is arbitrary, so there is the need to transform the original solution to a rotated solution that has desirable properties concerning interpretation (Kaiser, 1958).

Maximum likelihood factor analysis assumes that the variables have a multivariate normal distribution. If the variables contain a substantial amount of measurement error, which is often the case in the behavioral and social sciences (Dunteman, 1989), then factor analysis has an advantage over principal components analysis. The common factors are uncontaminated by measurement error because measurement error is part of the unique variance which is uncorrelated with the common variance, whereas principal components contain measurement error (Jolliffe, 1986).

Correspondence analysis deals with categorical variables that are crossed to define cells for which frequencies are available from a sample of observations (Greenacre, 1984). The rows of the contingency table are treated as observations and the columns of the table as variables; on this, a principal components analysis of a transformation of the contingency table is performed, and principal componenent score vectors generated for the rows (Jolliffe, 1986). After transformations of these score vectors, the scores associated with the largest two principal components are plotted in two or three dimensions so that clusters can be identified and correspondences between the row and column variables can be visually identified. Correspondence analysis is most useful when both the number of row and column categories are large (Dunteman, 1989).

In correspondence analysis, a crosstabulation table of frequencies is standardized, so that the relative frequencies across all cells sum to unity. In the table of relative frequencies, the observations are represented in terms of the distances between individual rows and/or columns in a low-dimensional space. The computation of the relative frequencies is distributed as one unit of mass across all cells in the contingency table. The row and column totals of the matrix of relative frequencies are called the row mass and column mass, respectively (see Greenacre, 1984).

Inertia in correspondence analysis is used by analogy with the definition in applied mathematics of "moment of inertia," which stands for the integral of mass times the squared distance to the centroid (Greenacre, 1984). Centroids, or weighted averages of a set of points are displayed in multi-dimensional space. Distances between profiles are measured by χ^2 distance, a variant of the Euclidean distance function, where factors are weighted by the squared difference term as an additional factor in the denominator of each squared term of the Euclidean distance (for a simple explanation of correspondence analysis, see Clausen (1998), for a more in-depth detailed explanation, Greenacre, 2007). Since the masses add up to 1, the inertia is the weighted average of the squared χ^2 distances between row profiles and their average profile. If the rows and columns in a contingency table are completely independent of each other, the entries in the table (distribution of mass) can be reproduced from the row and column totals alone, or "row / column profiles" in the terminology of correspondence analysis. The inertia is high when the row profiles have large deviations from their average, and are low when they are close to the average (Greenacre, 2007).

4.4 Results

4.4.1 Regression Analysis

The initial exploratory multiple regression analysis was performed by regressing the log-OCI on the four questionnaire items Importance, Knowledge, Prediction and Luck, as well as the portfolio values return estimate, realized return, mean diversification and standard deviation of returns.

The regression results in Table 4.1 have an adjusted R^2 of 0.37, and the F(8,331)-

Table 4.1: Regression Results of log-OCI on all Variables

Coefficient	Estimate	Std. Error	t-Value	p-Value
(Intercept)	4.89	0.95	5.11***	< 0.001
Importance	0.13	0.20	0.68	0.496
Knowledge	−0.15	0.25	−0.62	0.533
Prediction	0.30	0.22	1.41	0.160
Luck	0.29	0.17	1.73*	0.086
Return Estimate	0.01	0.00	5.48***	< 0.001
realized return	−0.88	0.08	−10.98***	< 0.001
Mean Diversification	−7.82	0.94	−8.27***	< 0.001
Return Stdev	−0.014	0.01	−1.05	0.293
R^2			0.37	
F-Stat			26.6***	< 0.001
DW-Stat			2.0	0.978

Note: The top panel in the Table shows the regression results from the full unrestricted model regression in Equation 4.1. The rows are labeled questionnaire values Importance, Knowledge, Prediction, and Luck. These values are ex-ante answers derived from the psychological questionnaire administered before trading commenced. The values in the middle panel are classical portfolio values calculated ex-post after trading had been completed, based on similar studies, e.g. by Odean (1998a,b).
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

statistic of 26.6 has a p-value < 0.001. The Durbin-Watson test for autocorrelation in the residuals does not confirm autocorrelation (Durbin-Watson statistic = 2.0, p-Value = 0.978).

In order to test for evidence of multicollinearity, which would render information within the regression redundant, we calculate a correlation matrix containing all cross-correlations between the regression variables.

Visual inspection of the correlation matrix in Table 4.2 indicates high correlations (0.6 to 0.8) within the values from the questionnaires in the top left area (see also panel 1 in Table 4.1), and low to zero correlations between questionnaire items and portfolio variables (see bottom left area values). Amongst the portfolio variables, the main correlation is a (weak) positive correlation of 0.26 between log-overconfidence and the return estimates, and a negative correlation of -0.43 between realized return and log-overconfidence (see bottom-right area of Table 4.2). This is to be expected, as the higher the overconfidence value, the lower the return. The relationships between ex-ante questionnaire answers and ex-post portfolio variables, as well as amongst these two categories themselves is explored in more detail in the following factor analysis (see Table 4.5).

In order to confirm or reject multicollinearity within the correlation matrix in Table 4.2, we calculate the Variance Inflation Factor (VIF), which is shown for all the regres-

Table 4.2: Correlation Matrix

	Import.	Knowl.	Predic.	Luck	Ret.Est.	Real.Ret.	log-OCI	Mean DoD	Stdev
Import.	1	0.80	0.60	0.06	0.12	0.06	0.10	−0.09	0.01
Knowl.	0.80	1	0.73	0.01	0.13	0.10	0.09	−0.14	0.00
Predic.	0.60	0.73	1	0.06	0.18	0.06	0.14	−0.09	−0.08
Luck	0.06	0.01	0.06	1	−0.03	0.02	0.06	−0.01	0.13
Ret.Est.	0.12	0.13	0.18	−0.03	1	−0.03	0.26	0.02	−0.08
Real.Ret.	0.06	0.10	0.06	0.02	−0.03	1	−0.43	−0.14	−0.03
log-OCI	0.10	0.09	0.14	0.06	0.26	−0.43	1	−0.29	−0.01
Mean DoD	−0.09	−0.14	−0.09	−0.01	0.02	−0.14	−0.29	1	−0.11
Stdev	0.01	0.00	−0.08	0.13	−0.08	−0.03	−0.01	−0.11	1
VIF	2.76	3.83	2.21	1.03	1.04	1.03		1.05	1.05

Note: The Table above shows the correlations between each of the regression parameters from the regression in Equation 4.1. The first four columns/rows are labeled questionnaire values Importance, Knowledge, Prediction, and Luck. The following columns/rows are labeled the Return Estimate, Realized Return, the regressor log-OCI, the Mean Degree of Diversification, and the Standard deviation of returns. The VIF (Variance Inflation Factor) in the last line is calculated for all the regressors. The high correlation values in the cells Importance, Knowledge, and Prediction on the one hand, and the values in the cells Realized Return, log-OCI, and Mean DoD relate to the factors extracted in the factor analysis below, shown in the two panels separately in Table 4.5.

sors in the last line of Table 4.2. A value of greater than 5 would indicate collinearity (Heiberger and Holland, 2004), and as one can see, none of the values come close, so we can reject the implicit null hypothesis of collinearity within the data[15].

In previous literature, either authors have analyzed overconfidence via observable ex-post portfolio performance and structure factors, such as Barber and Odean (2000), Barberis and Huang (2001), or Eckholm and Pasternack (2007), or have administered ex-ante questionnaires (see e.g. Glaser et al., 2007) and derived results on overconfidence from comparison of the questionnaire results and following subject behavior. To the best of the author's knowledge, no other extant research has attempted to determine potential reasons for overconfidence in trading behavior from both ex-ante data as well as from ex-post analysis of portfolio development over time. In our current research, in order to find determining factors for the development of overconfidence, we analyze both categories of data - ex-ante questionnaire answers, and ex-post portfolio results.

The regression output shown in the middle panel of Table 4.1 and Table 4.2 provide an indication that the interrelations between questionnaire answers and portfolio calculations seem to be contained within each category rather than between categories, so

[15]The variance inflation factor (VIF) quantifies the severity of multicollinearity in an ordinary least squares regression analysis. It provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. A cutoff factor of 5 is often used (Sheather, 2009).

we analyze each data category separately, i.e. overconfidence as potentially explained by either ex-post portfolio calculations on the one hand, and in relation to ex-ante questionnaire answers, on the other hand.

The parameters contained in the middle panel of Table 4.1 show the relationship between overconfidence and return estimate, return achievements, mean diversification of portfolios and the riskiness, i.e. standard deviations of realized returns. In Chapter 3, these relationships between the portfolio values and resulting over- or underconfidence have already been analyzed. The results in the middle panel of Table 4.1 confirm the previous results that, a) on average, the higher the return estimate, the more likely the subject to be overconfident, b) that the higher the realized return, the lower the resulting overconfidence, c) that the more diversified the portfolio, the less overconfident the participant turns out to be, and d) that the riskiness of the portfolio is not related to subsequent overconfidence (see Chapter 3).

However, despite this, the results are not totally satisfactory, as the highly significant coefficients from the regression - the forecasted return, the realized return and the standard deviation of returns are those values that are expected to correlate strongly negatively (in the case of the realized return, the mean diversification, and the standard deviation of returns), or positively (in the case of the return estimate) with the log-OCI (see Table 4.1).

In order to shed more light on the potential predictors of overconfidence rather than the statistical analysis of trading itself, we perform a multiple regression of the log-OCI on just the questionnaire results as a restricted model, regressing for each participant i the log-OCI on the four questionnaire items described as Q_j in the following multiple regression:

$$logOCI_i = \alpha + \beta_j Q_{ji} + \epsilon_i \tag{4.2}$$

The results from this regression are tabulated in Table 4.3.

Table 4.3 shows the results of the regression of log-overconfidence on the aggregated answers to the questionnaire. Even though the adjusted R^2 is only .01, one of the answers shows significant (at about the 5% level) correlation with overconfidence. The self-declared ability of participants to be able to forecast outcomes of events seems to be linked to subsequent underperformance due to overconfidence. This is an indication that is worth exploring in more detail, as the multiple regression results provide some indication of the origin of overconfidence in the data set, but does not provide enough of a cause-and-effect relationship for forecasting purposes.

Table 4.3: Regression Results of log-OCI on Answers

Coefficient	Estimate	Std. Error	t-Value	p-Value
(Intercept)	−0.43	0.96	−0.46	0.648
Importance	0.15	0.25	0.61	0.542
Knowledge	−0.19	0.31	−0.63	0.532
Prediction	0.51	0.27	1.91*	0.058
Luck	0.19	0.21	0.92	0.356
adj. R^2			0.01	
F-Stat			2.02*	0.091
DW-Stat			2.01	0.876

Note: The table above shows the regression results from the reduced model regression in Equation 4.2. The rows are labeled questionnaire values Importance, Knowledge, Prediction, and Luck.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.4.2 Exploratory Factor Analysis

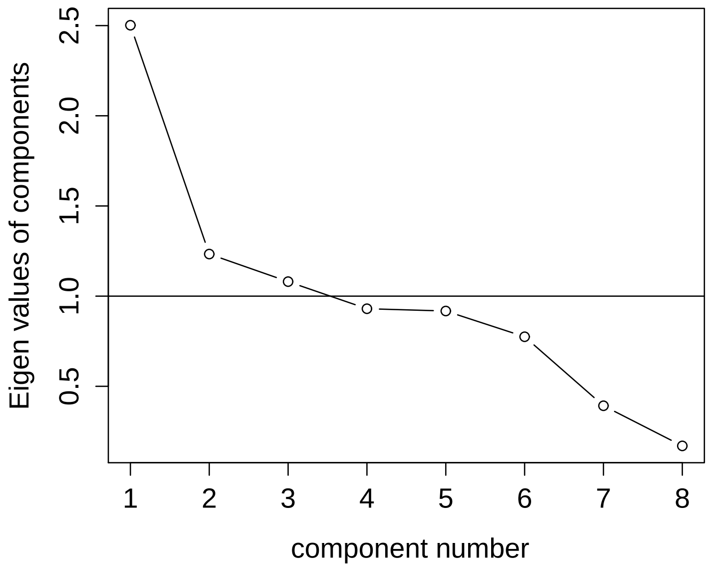
As the regression analysis provides only a slim indication of where to look for determinants of overconfidence, factor analysis could potentially shed more light on underlying relationships rather than just relying on direct linear relationships in a (multiple) regression analysis. Factor analysis has the advantage of reducing the number of explanatory variables to a minimum that can potentially explain commonalities in the data set.

The factor analysis performed in this analysis includes all 8 questionnaire and portfolio result data categories as outlined above in the full model regression analysis (compare Equation 4.1) with the log-oci as the explained variable, in order to identify a reduced set containing potential "underlying" variables concatenated into common factors depending on their common correlations.

As discussed in the Measures section, we use a scree plot to identify a potential initial cutoff point for the maximum number of factors to extract. The scree plot plots the eigenvalues against the possible factor numbers. The point where the slope of connected eigenvalues begins to level off from the steeper initial slope is generally taken to be the cutoff point for the maximum number of factors to extract. Another popular method is to use those factors whose eigenvalues are greater than one (see e.g. Kim and Mueller, 1978a, pp. 43-45).

The scree plot in Figure 4.1 shows that the optimal number of factors to extract should be set to no more than 3-4 (3 if one chooses the eigenvalue criterion, 4 if one chooses where the slope of the line between eigenvalues begins to level off). However, after running the factor analysis 4 times, each time varying the number of factors

Figure 4.1: Factor Analysis Scree Plot



from 1 through 4, the Bayesian Information Criterion (BIC) reveals that the optimum number of factors is actually 1 (BIC = -80.61, the highest negative value in column BIC of Table 4.4). The χ^2 test of the hypothesis that the given number of factors is sufficient confirms that 2 factors are sufficient to explain the data ($p < 0.016$ and a factor reliability of 0.964).

Table 4.4: Bayesian Information Criterion Table

Factor	BIC	χ^2	p-Value
1	**−81.61**	35.97**	0.016
2	−57.07	18.71	0.130
3	−31.36	9.44	0.220
4	−9.88	1.78	0.410

*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Results from the factor analysis are tabulated in Table 4.5. The first four rows contain the questionnaire items; the remaining rows comprise portfolio calculation variables.

The fit based upon off-diagonal values is 0.96 (indicating that the predicted off-diagonal covariance matrix elements explain 96% of the observed covariance, see e.g. Kim and Mueller, 1978a, or Kim and Mueller, 1978b), and score adequacy (correlation of factors with oberved variables) is high, with correlation of scores (observed variables) with factors at 0.96 (ML1). The multiple R^2 of the observed variables with the factors is 0.97 (ML1). For detailed explanations of interpretation of factor analysis, see Grice (2001), or Gorsuch (1997).

Analysis of the factors reveals that the factor ML1 has high positive scores in three questionnaire values - the importance of soccer, perceived knowledge of the subject matter and purported prediction ability (all shown in boldface in Table 4.5). It is interesting that the final remaining questionnaire coefficient, the perceived impact of luck on game outcomes, exhibits almost no factor loading. The ML1 factor can therefore be interpreted as a perceived ability factor based on participants' estimation of their own prowess concerning their knowledge and prediction abilities given that the topic (soccer) is of importance to them.

Table 4.5: Factor Analysis Results

Coefficient	ML1	h^2	u^2
Importance	**0.81**	0.66	0.34
Knowledge	**0.98**	0.96	0.04
Prediction	**0.74**	0.55	0.45
Luck	0.02	0.00	1.00
Return Estimate	0.14	0.02	0.98
Realized Return	0.10	0.01	0.99
Mean Diversification	−0.14	0.02	0.98
Stdev of Returns	−0.01	0.00	1.00
R^2	0.97		

Note: The values in the column ML1 are the maximum likelihood factor loadings that best explain the observed correlations amongst the observed data. The column h^2 contains the communality; the proportion of the variance of an item that is accounted for by the common factors. The column u^2 contains the unique variance of the respective observed variable; it is given by 1 - h^2 - the specific variance plus the error variance or random error (see Kim and Mueller, 1978a,b).
The first panel contain the factor analysis for the four questionnaire answers (ex-ante values, as they are collected before trading begins), the following panel contains the corresponding values for the portfolio values (ex-post values, as they are evaluated after trading has ended). The bottom panel shows the R^2 for the ML1 factor. The high factor loadings in the top panel correspond to the correlations in Table 4.2 in the top left area (0.6 to 0.8). The factor loadings in the middle panel correspond to the portfolio value correlations in the bottom right area of Table 4.2 (-0.14 to -0.43).

The calculation results indicate that only three variables have any predictive power. The factors Importance of the subject matter, Knowledge of the subject matter, and perceived Prediction abilities are all related and constitute the main explanatory variables. The questionnaire item Luck in predicting soccer outcomes does not explain overconfidence well, and as already diagnosed, the portfolio values in the second panel of Table 4.5 are also not good explanatory factors for overconfidence. The fact that the factor analysis yields the result that return estimates are positively correlated and diversification negatively correlated with overconfidence are also not surprising, due to the fact that the less diversified a participant's portfolio is, the more potential the participant has for large losses, meaning that this participant is likely to be overconfident.

All in all, the factor analysis results are therefore not entirely unexpected, insofar as the calculated portfolio values are not good predictors of overconfidence. However, the strong relationship in the first three questionnaire answers forms the basis for further analysis, which may shed light on potential previously unidentified correlations in the current data set between attitudes and overconfident trading results.

4.4.3 Exploratory Correspondence Analysis

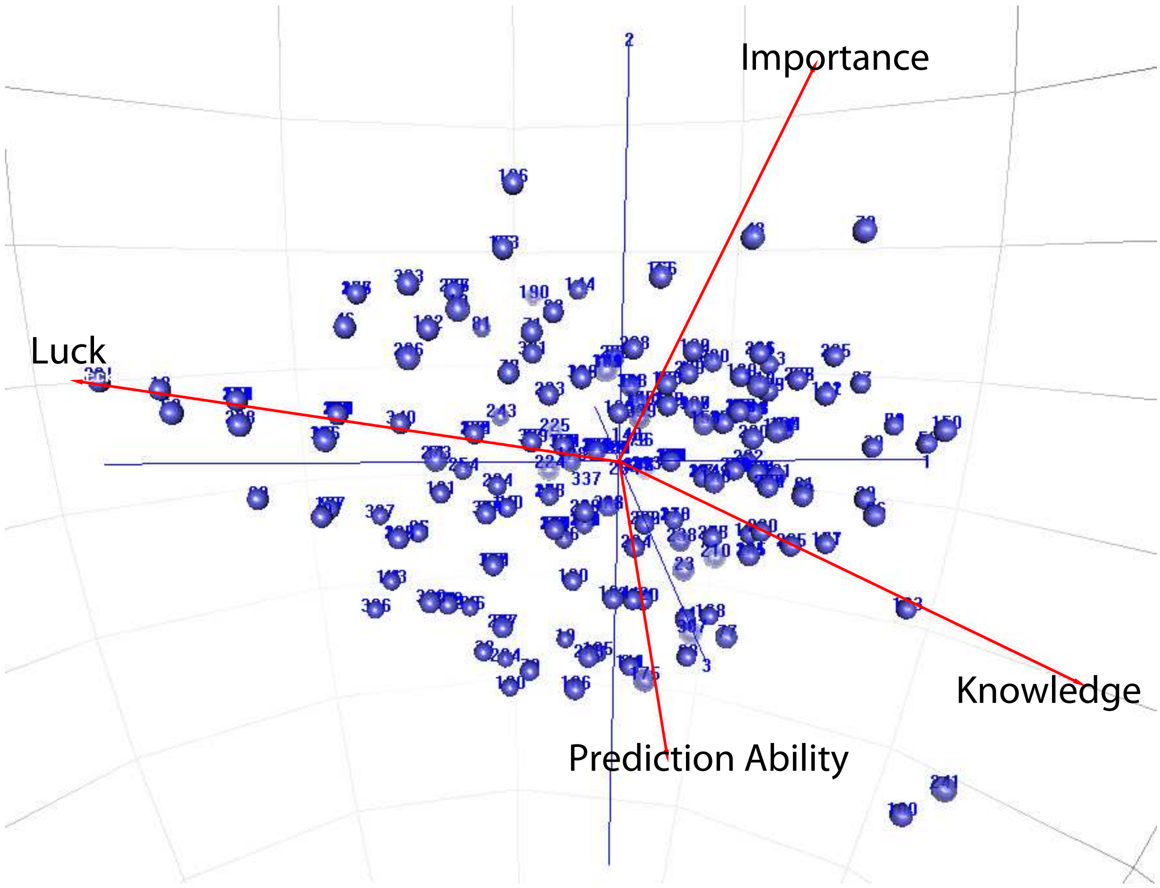
After having established that explanations for overconfidence in the data set cannot be concentrated exclusively on the portfolio analysis variables (from the regression analysis), but should rather be focused additionally on the three questionnaire items Importance, Knowledge, and Prediction as mentioned above, resulting from the factor analysis results. The importance of the items from the questionnaires posed ex-ante form an important guide to the intentions of the participants in their trading actions, which are measured ex-post by looking at the realized results (via the portfolio calculation values).

In order to find relationships that enable us to attempt to forecast overconfidence, we use correspondence analysis on the questionnaire items to find clusters within the data with similar characteristics.

Narrowing down and focusing on the relationship between the questionnaire answers and overconfidence, we perform an exploratory correspondence analysis with the log-OCI as the dependent variable and the 4 questionnaire answers as the explanatory variables. Correspondence analysis establishes a multidimensional representation of the association between the row and column categories of a two- or multi-way contingency table, and can be visualized in two or three dimensions as a reduced-rank approximation to the data set in question. We visualize the four categories in a three-dimensional graph, shown in Figure 4.2. The axes correspond to the factor loading combinations of the 4 coefficients (answers to the questions Importance and Knowledge of subject matter, importance of Luck and estimated Prediction ability).

The mass points in a three-dimensional depiction of the correspondence analysis are arranged by their χ^2 distances from the principal axes. The overall χ^2 statistic identifies a small number of dimensions in which the deviations from the expected values can be represented (This is similar to the goal of Factor Analysis, where the total variance is decomposed, in order to arrive at a lower-dimensional representation of the variables representing most of the covariance matrix of variables). The principal

Figure 4.2: Correspondence Analysis 3D Mass Points



Note: The Figure above shows in three-dimensional space the factor loading combinations of the 4 coefficients (answers to the questions Importance and Knowledge of subject matter, importance of Luck and estimated Prediction ability). Each of the 340 mass points in the graph is associated with the individual factor loading of each of the 340 participants. On the principal axes, the mass points are distributed according to their χ^2 distances from the mean (the origin).

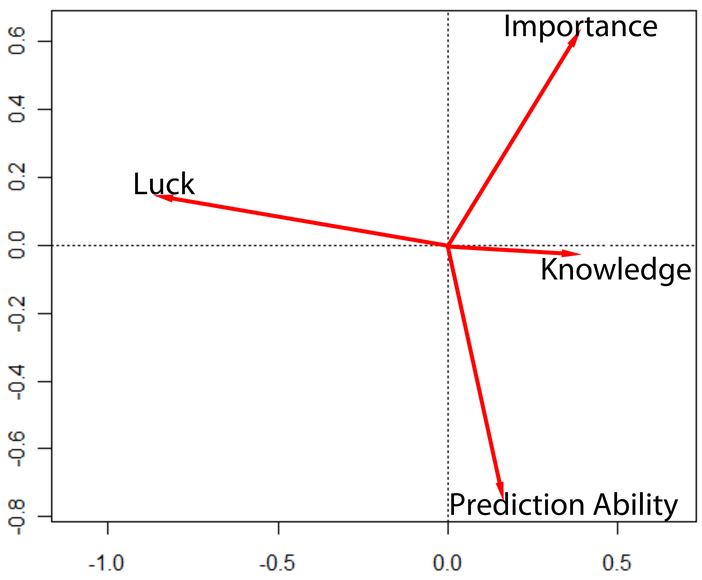
inertia is determined as χ^2 / total N (340 in our case). Each value is then arranged along the principal axes by its individual contribution to the principal inertia.

The principal axes are the straight (red, in Figure 4.2) lines that come closest to the profile points in the sense of least squares (an explanation of the profile points is provided in the Methods Chapter, Subsection Exploratory Correspondence Analysis). An axis passes through the row centroid (weighted average distance of an element in a row of the contingency table) with is at the origin of the axis. All row profiles are projected onto this axis. The first principal inertia is then the weighted sum of squared distances from these projections to the centroid. This is then performed for the further principal inertias on the other axes.

The axes without mass points are displayed in two dimensions in Figure 4.3 for clarity. The distances on the axes correspond to χ^2 distances from the centroid (origin).

Each mass point shown in Figure 4.2 corresponds to the location of a participant according to the loading on each item to the questionnaire, depending on how the participant answered on the 6-part scale. So, e.g. a participant who answered 6 (agree strongly) on Importance and 6 (agree strongly) on Knowledge could be found close to the top, front and right in a three-dimensional cube.

Figure 4.3: Correspondence Axes



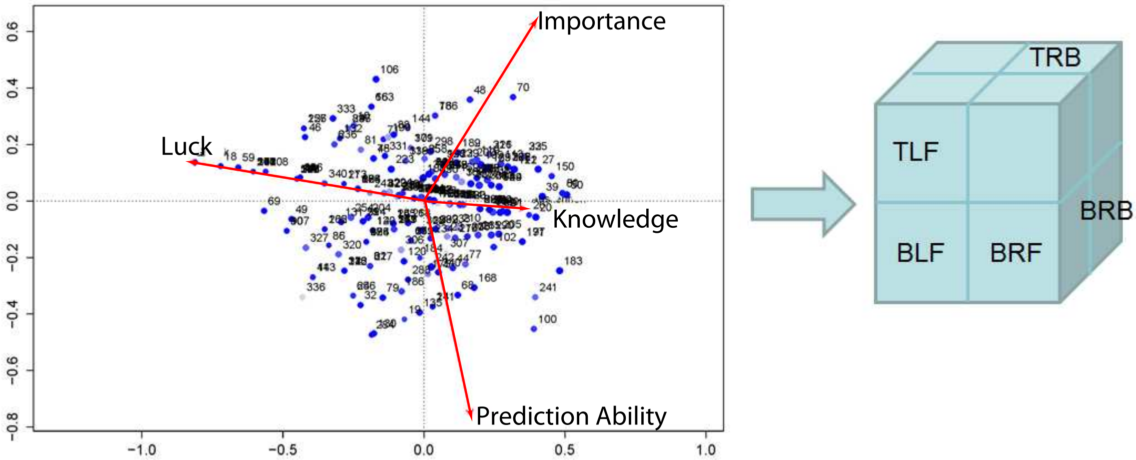
Note: The Figure above shows in two-dimensional space the least-squares lines (these are then planes in three-dimensional space) corresponding to each of the 4 questionnaire answers. The mass points of each observation (340 participants) would then be placed according to the respective χ^2 distances around each of the lines. The axes correspond to the principal axes from the decomposition in the correspondence analysis, and the axis numbering is in χ^2 distance.

We analyze the points contained in the 3-dimensional space and separate this space into 8 quadrants of a three-dimensional cube (front/back, top/bottom, left/right, see Figure 4.4). In order to identify corresponding loadings as independent variables on overconfidence as the dependent variable, we assign each participant to one of the 8 quadrants as identified by the mass points shown in Figure 4.2.

This in effect means that a participant is assigned a quadrant based on the level of their answers to the questions from the questionnaire. Next, the participants assigned to each quadrant are evaluated according to their level of overconfidence (measured by the log-OCI) and the sum of (just the) overconfident participants in each quadrant is taken. The results are shown in Table 4.6.

From the sum of overconfident participants in relation to the total number of participants in each quadrant, we calculate the percentage of overconfident participants contained in each quadrant versus all participants in a quadrant. The rightmost column

Figure 4.4: Correspondence Quadrants



Note: The Figure above shows in two-dimensional space on the left the (three-dimensional) correspondence analysis, and how the three dimensions are separated in to a cube containing 8 quadrants. The individual quadrants are labeled according to their location, e.g. TLF = Top Left Front quadrant, BRB = Bottom Right Back quadrant, etc.

in Table 4.6 shows the questionnaire item that corresponds to the respective quadrant. Of interest are the values in bold face. In the Top Right Back quadrant (TRB), 72% of participants whose mass points are located in this quadrant have proven to be overconfident, and 76% of the mass points (one per participant) in quadrant Bottom Right Back (BRB) are overconfident.

Table 4.6: Correspondence Analysis Summary

Quadrant	Number of OC Participants	Total Number of Participants	% overconf.	Questionnaire Item
TopLeftBack TLB	34	55	62%	
TopRightBack TRB	28	39	**72%**	**Importance**
TopLeftFront TLF	34	60	57%	Luck
TopRightFront TRF	21	38	55%	
BottomLeftBack BLB	24	40	60%	
BottomRightBack BRB	29	38	**76%**	**Prediction**
BottomLeftFront BLF	14	31	45%	
BottomRightFront BRF	24	39	62%	Knowledge
Sum	208	340		

Note: Participants are categorized by their level of overconfidence as measured by an overconfidence index (OCI) value of greater than 1.

As the final stage in our analysis, we consider only the factors with the highest loadings. The highest loadings from the correspondence analysis is on the two factors "Importance" and "Prediction" (see the boldface items in Table 4.6), and we analyze participants' answers to these two questions. In Table 4.7, the answers to the two questionnaire items are shown individually and taken together, and the number of participants who proved to be overconfident are contrasted to the total number of participants whose answers were > 3 for the questionnaire items.

Table 4.7: Answers to Questions Importance and Prediction

Question	Total No. of Participants	OC No. of Participants	% overconf.
Importance > 3	111	79	71.2%
Prediction > 3	117	83	70.9%
Imp. and Predic. > 3	74	56	75.7%

Note: The Table above shows how many participants answered > 3, i.e. 4, 5 or 6 (”agree somewhat”, ”agree” or ”agree strongly”) on the 6-point Likert scale for the questionnaire items Importance, Predicion and the combination of both. In the last line, Importance is shortened to “Imp.”, and Prediction is shortened to “Pred.”. The right-most column shows the percentage of those participants who answered in the affirmative to the questionnaire items who later proved to be overconfident in their trading. The Table shows ex-ante values, i.e. the answers to questionnaire items that were answered before trading commenced (in contrast to ex-post overconfidence values such as in Table 4.6).

Table 4.7 shows who answered that the topic was ”slightly more important”, ”more important” or ”much more important” and/or that their prediction abilities were ”slightly better”, ”better” or ”much better” than others. This corresponds to the values 4,5 or 6 (or ”agree somewhat”, ”agree” or ”agree strongly”) on the 6-point Likert scale. In Table 4.7 one can see that although 71% of candidates that scored 4-6 each individually on the items Importance or Prediction, 76% of those who scored 4-6 on both of these items simultaneously prove to be overconfident.

It is important to separate the values that were determined ex-ante - these are the values derived from answers to the questionnaire items provided to participants before trading commenced. These values are contrasted with ex-post variables such as the level of overconfidence as measured by the (log) Overconfidence Index. These values are calculated after trading was completed. The combination of ex-ante and ex-post variables provides the basis for our analyses into the reasons behind eventual overconfident behavior in trading.

4.5 Discussion

The four steps of the analysis outlined in the Methods section were performed in order to identify the drivers of overconfidence and to indicate a possible forecasting methodology as an early warning signal for risk managers.

In the multiple regression analysis performed first, we attempt to isolate the factors that correlate highly with the observed ex-post overconfidence as measured with the log-overconfidence index. Even though significance and F-test results were quite high, the factors that correlate strongly with log-overconfidence do not provide enough information for forecasting purposes.

The higher the return estimate, the higher the chance that one could be overconfident, this is to be expected, as it results directly from overconfidence as defined e.g. by Lichtenstein et al. (1982) in the form of miscalibration (for the definitions of overconfidence in previous literature, see also Chapter 3, Overconfidence in Soccer). The mean degree of diversification is strongly negatively correlated with the log-OCI, indicating that the more overconfident the investor, the less diversified he is - no risk, no fun. High diversification in this sense is a sign of prudence, not of excessive risk-taking behavior. The lower the realized return, the higher the degree of overconfidence - this is also to be expected, as it conforms to results from previous research e.g. by Barber and Odean (2000) or Odean (1998b) - these results are therefore not surprising.

As the variables Return, Diversification and Risk (measured as the standard deviation of returns) do not provide additional information for forecasting purposes, the multiple regression on just the answers to the questionnaire sheds more light on the relationship between participants' initial judgments versus the ex-post overconfidence. Here a weak (linear) relationship with individuals' self-declared ability to predict game outcomes is observable, but the linear relationship is not sufficient to warrant using this variable as a sole linear predictor of overconfidence.

The factor analysis reveals that one common factor underlies the result structure, the factor (ML1) showing very high loadings on three of the four questionnaire answers which is interesting - this factor represents convictions concerning personal ability. The questionnaire item concerning whether a participant believes that luck plays a role in determining the outcome of games has a loading of nearly zero, whereas the three items Importance of Soccer, Soccer Knowledge and Prediction Ability have extremely high loadings. The ex-post portfolio values (realized return, portfolio diversification and the standard deviation of returns) as well as the ex-ante return estimate exhibit very low factor loadings within the factor ML1, indicating, that they do not add explanatory power to the question of what drives overconfident behavior in trading.

Based on the results that professed personal ability seems to be an important driver in future overconfidence, a correspondence analysis reveals that the two most important drivers of overconfidence are the questionnaire items concerning the importance of soccer for the participant and the professed ability to be able to predict the outcome of games. The correspondence analysis shows the mass points as loadings in three-dimensional space (see Figure 4.2). If one imagines the space spanned by the axes as divided into 8 quadrants, those quadrants containing the mass points for Soccer Importance and Prediction Ability are the top right back quadrant (quadrant TRB) and the bottom right back quadrant (quadrant BRB), respectively (see Figure 4.4. Of all participants' mass points located in these two quadrants, 72% of those located in the first selected quadrant prove to be overconfident participants, and 76% of the second quadrant are overconfident (see Table 4.6). In a last step, we combine the selected participants from the two quadrants with the values of their answers to the questionnaire items Soccer Importance and Prediction Ability.

We analyze those participants who answered positively (i.e. 4, 5, or 6 on the 6-point Likert scale) to both the questions Soccer Importance and Prediction Ability individually and taken together. Combining positive answers on both questionnaire items, 56 of the identified 74 participants (i.e. 76%) proved to be overconfident. Additionally, of the top 20 most overconfident participants, 10 are contained in this sample.

This identifies those participants who will prove to be overconfident before trading has started from the quality of selected answers to the questionnaire items. If one takes those participants after having completed the questionnaire who answered 4, 5, or 6 on both of the items concerning how important the subject matter is to them and how much they believe in their own prediction ability, one is able to say that within this sample, roughly 76% of them will prove to be overconfident in their trading results three weeks later, after the market has closed.

4.6 Robustness of Overconfidence Results

In order to verify that the above results are not due entirely to chance, we analyze the robustness of the results. There are different techniques for robustness analysis or validation. For example, in data mining, validation of a model involves splitting the data sample, e.g. into two unequal sized partitions, and using the second partition of data to validate the results achieved from the first partition (see e.g. Bishop, 2006, or Witten et al., 2011; see these authors also for detailed explanations of additional methods such as cross-validation and bootstrapping). According to Cohen and Cohen (2008), *"Validation refers to the process of applying the model to data for which the model had not been fitted. Ideally, one would like to fit the regression to data from a population and then validate it for data from another population. Short of this, we can simply exclude part of the data, fit the model and then validate it on the excluded* *data"* (Cohen and Cohen, 2008, p. 536). This can be performed longitudinally, i.e. partitioning the data into subsamples and re-estimating, e.g. the first 75% versus the last 25% of all trades.

Another method of validating robustness would be by deriving the inverse relationships from the observed results and confirming these empirically from the data, or by analytically examining the data, e.g. using Bayes' theorem. The technique of inverse validation is described by Liu and Labrosse (2002), and in a Bayesian context by Bhattacharya and Haslett (2007). We follow these two techniques for validating the robustness of our results.

4.6.1 Empirical Derivation of Inverse Relationships

Our results state that overconfidence can be inferred from slight to strong concurrent agreement with the questionnaire items Importance and Prediction. The inverse would imply that if participants answered that they disagreed (score 1-3) with the items Importance and Prediction that this would imply underconfidence. In order to determine whether this is the case, we recalculate the analog ex-post underconfidence values from the overconfidence Table 4.6, the corresponding underconfidence values are shown in Table 4.8.

Working backwards, we identify the ex-post underconfident participants from the mass points in the quadrants identified in the previously performed correspondence analysis. From the positioning of these mass points within the 8 quadrants in three-dimensional space (shown previously for overconfident values in Figure 4.2), we identify those quadrants containing the axes aligning Importance, Knowledge, Prediction Ability, and Luck and determine whether the mass points indicating underconfidence lie within those quadrants that do not contain significant numbers of mass points from the overconfident participants (the inverse to the detection procedure for overconfidence shown in Table 4.6).

The values from Table 4.8 depicting underconfidence per quadrant (as was performed above in Figure 4.4 and Table 4.6), however, do not show anything like the significance of the corresponding overconfidence values from Table 4.6. In actual fact, the highest value, 55% underconfidence in a particular quadrant (shown in bold face in Table 4.8), refers to the quadrant bottom/left/front, which does not have any major correspondence with any of the factor loadings on the questionnaire items. Detailed analysis of the mass points in this quadrant do not show any correspondence with

Table 4.8: Underconfidence per Quadrant

Quadrant	Number of UC Participants	Total Number of Participants	% underconf.	Questionnaire Item
TopLeftBack	21	55	38%	
TopRightBack	11	39	28%	Importance
TopLeftFront	26	60	43%	Luck
TopRightFront	17	38	45%	
BottomLeftBack	16	40	40%	
BottomRightBack	9	38	24%	Prediction
BottomLeftFront	17	31	**55%**	
BottomRightFront	15	39	38%	Knowledge
Sum	132	340		

Note: The Table above shows ex-post underconfidence per quadrant analogously to Table 4.6. The percentage in boldface shows the largest percentage of ex-post underconfident participants in any one quadrant.

the factors Importance, Luck, Prediction or Knowledge. The corresponding values are tabulated in Table 4.9.

Table 4.9: Underconfidence vs. Importance and Prediction

Question	Total No. of Participants	UC No. of Participants	% underconf.
Importance < 4	229	100	44%
Prediction < 4	223	98	44%
Imp. and Predic. < 4	186	84	45%

Note: The Table above shows how many participants answered < 4, i.e. 1, 2 or 3 ("disagree strongly", "disagree somewhat" or "disagree slightly") on the 6-point Likert scale for the questionnaire items Importance, Predicion and the combination of both. The right-most column shows the percentage of those participants who answered in the affirmative to the questionnaire items who later proved to be underconfident in their trading.

From Table 4.8, a brief inverse check of the overconfident participants who "disagreed" to both Importance and Prediction reveals that 102 of 186, or 55% of the mass points are contained in this subset, neither values providing insight into potential sources of underconfidence. Table 4.9 shows ex-post numbers of underconfident participants and what these participants answered ex-ante to the questionnaire items Importance and Prediction separately, and both items in combination. Of those participants who ex-ante answered both that they did not attribute importance to the subject matter and also that they did not consider themselves able to predict the outcomes of soccer matches, only 45% turned out to be ex-post underconfident - not a particularly good ex-ante predictor for ex-post underconfidence.

What could then constitute further appealing combinations indicating underconfidence? One might think that a high agreement with the questionnaire item Luck could indicate a tendency to underconfidence, but brief analysis of intuitively appealing combinations from Table 4.6 (those quadrants that do not contain concentrations of mass points indicating overconfidence, such as e.g. quadrant fop/left/front, containing the axis Luck), show that correspondence of underconfidence with Luck is 16%. Intuitively, participants who indicate that they know a lot about soccer might prove to be underconfident due to the fact that they purport to be very knowledgable about the subject matter, and therefore more cautious, but this is also dispelled by analyzing the data (only 37% of the participants who answered that they were knowledgeable about soccer proved to be underconfident).

In short, it is much easier to find proof of presence of overconfidence rather than of its absence. None of the inverse relationships nor the intuitively appealing combinations of underconfidence with questionnaire items reveal substantive statistically relevant relationships. Conversely, the absence of such inverse relationships strengthen the intuitive reasoning that the results concerning overconfidence are robust rather than pure artifacts.

Analytical Bayesian Robustness

A final look at robustness of the reported results consists of an analysis of the probability that the reported results are due only to chance. The probability of those participants who responded strongly to both questionnaire items that have proved to provide the most insight into eventual overconfidence (Importance of subject matter and self-professed ability to Predict outcomes) then ending with overconfident results occurring just by chance is quite small. If one posits that the chance of either disagreeing (i.e. selecting answers 1-3) or agreeing (i.e. selecting 4-6) on each of the questions is 50% and that answers to each of the questions are independent (for argument's sake), then the probability of agreeing on two independent items concurrently is 0.50*0.50 = 0.25. Even if one uses the posterior probability that a randomly selected participant will finish overconfident (208/340 = 61.2%) rather than equal probability, the chance of a participant finishing overconfident conditional on having agreed to two of the four questionnaire items is, according to Bayes' Theorem, still only 61.2%. Under the assumption that over- and underconfidence are equally likely, the result is 50%. Either way, our findings that 76% of participants link overconfidence with combined agreement on the questionnaire items Soccer Importance and Prediction Ability are unlikely to have occurred just by chance.

4.7 Conclusion

In this paper we analyze how overconfidence can be forecasted, based on data from an experimental futures market, described in the Data subsection of the Method section (section 4.2.2). Previous research does not consider the topic of how to forecast overconfidence with a view to limiting it by risk managers in an institutional trading or investment environment. We perform a multi-stage analysis consisting of a multiple regression analysis of log-overconfidence (measured by the log-overconfidence index defined in a previous chapter) on 8 variables, 4 of which are from the psychological questionnaire provided to participants before trading commenced, the remaining 4 variables comprising traditional portfolio statistics (ex-post values revealed after trading).

The second step consists of a reduced multiple regression of log-overconfidence on just the 4 questionnaire items, resulting in a weak connection with one of the items (Prediction Ability). Stage 3 was a factor analysis, one of the two resulting factors giving the indication that the three questionnaire items related to individual ability (Importance, Knowledge and Prediction) provide the highest loadings.

A correspondence analysis as the 4th stage and subsequent detailed analysis of the results show that overconfidence can be forecast by observing the two main contributing items from the questionnaire. Those participants who estimated that both (in combination) their professed ability to predict events and the importance of the subject matter determined their future overconfidence. 76% of our sample who scored highly in both of these categories subsequently lost substantial ground versus expectations in achieving investment returns.

These results prove to be robust, various analyses of different combinations of input variables and data cannot reject the previous findings.

This result is of particular relevance to risk management in institutional trading or investment environments. Much money is lost due to behavioral biases such as overconfidence. A suggestion is to periodically provide traders with psychological questionnaires focused on the two items previously isolated, e.g. in half-yearly review processes.

We suggest to provide psychologically determined questionnaire items focused on discovering traders with an increased sense of both importance of the particular topic compared with their peer group as well as on a heightened self-declared ability to predict events in the markets that objectively cannot be influenced by them.

With periodic results from well-defined questionnaire items, risk managemers would be in a much better position to determine which traders are susceptible to biasing their trading decisions with overconfidence, to monitor exactly those traders more carefully, to implement early-warning systems of overconfidence and to implement individualized stop-loss limits according to evidence of overconfident trading behavior.

With processes such as these in place, behavioral finance would be able to find its place in actively preventing biased behavior. Maybe this can be the beginning of behavioral risk management. As a result, pre-empting biased behavior with such early-warning systems based on behavioral finance techniques would contribute to markets becoming more efficient in the classical sense.

Chapter 5

Experimental and Real Dispositions

5.1 Introduction

The previous chapters have focused on modeling overconfidence as a miscalibration of expectations captured via questionnaire items versus actual behavior in trading situations on an experimental market. Overconfidence manifested itself as the difference between observed loss-making versus expectations of profit- or loss making. An effect similar to the effects of overconfidence can be observed as a consequence of a related cognitive bias, the so-called disposition effect. The disposition effect describes an investor's (often simultaneous) reluctance to close out losses and to hold on to gains (see e.g. Kaustia, 2010a). Siwar (2011) finds that overconfidence bias and the disposition effect are often linked in investment or trading settings.

In the previous experiment on the experimental prediction market described in the Data sections of Chapters 3 and 4, the questionnaire provided to participants before the commencement of trading asked participants to provide a forecast of their return achievement. Only 76 out of 340 in the final sample of participants (22%) estimated that they expected to make a loss. Slight overconfidence in the form of the better-than-average effect (Svenson, 1981) can be recognized insofar as 56% believed they would make a profit (the remainder estimated a zero-return outcome for themselves). The 209 participants (62%) who made a loss despite not believing it would happen to them were undoubtedly disappointed with their outcomes and would have preferred it to be otherwise.

Kahneman and Tversky (1979) found that losses cause people more severe pain (almost twice as much) than the pleasure derived from a gain of the same magnitude and called this phenomenon “loss aversion”. That people (not only professional traders and investors) are prone to a loss aversion bias was established originally by Kahneman and Tversky (1979) as a part of prospect theory. This was proposed by Kahneman and Tversky as a descriptive model of decision making under risk as an alternative to classical expected utility theory. Their basic hypothesis is that people evaluate potential losses and gains in their decisions rather than an absolute level of wealth.

Building on an article by Constantinides (1983) who discusses realization of gains and losses during the year versus at year-end, dependent on the U.S. tax code, Shefrin and Statman (1985) manifest this loss aversion bias as the disposition effect. They define it as investors holding on to loss-making positions for too long and selling profit-making positions too early.

The concept of maintaining profitable positions but limiting loss-making positions as soon as possible is not a new concept. David Ricardo (quoted in Kaustia, 2010a, p. 171) already invested according to his golden stock market rules in the late 1700's - his rules being "cut short your losses" and “let your profits run on” (Skousen, 2009). Shefrin and Statman describe a theoretical framework explaining investors' propensity not to follow Ricardo's advice (and personal experience as an extremely successful investor).

Even though Shefrin and Statman "framed" their paper in terms of December tax-loss selling, their seminal findings on the disposition effect are valid for a very wide range of situations and have been empirically tested under differing conditions. For empirical work on experimental markets, see e.g. Weber and Camerer (1998) or Haigh and List (2005). One early influential paper that measured the disposition effect on for-real markets was published by Odean (1998a), followed up by Barber et al. (2007) in addition to Grinblatt and Keloharju (2001). Odean's (1998a) methodology of measuring the disposition effect has become one of the de-facto standards for calculating the disposition effect, and is the methodology we follow. As outlined above, Shefrin and Statman (1985) identify four major elements that explain the disposition effect as a behavioral bias: 1) prospect theory, 2) mental accounting, 3) regret aversion and 4) self-control. The following sections provide an overview of these explanations of the disposition effect.

5.1.1 Prospect Theory

Prospect theory, developed by Kahneman and Tversky (1979) and Tversky and Kahneman (1974, 1992) was proposed as an alternative to conventional financial decision theory as formulated e.g. by Fama (1965), Malkiel (2007), or others, as subjects acting rationally via subjective expected utility theory on efficient markets. Instead, under prospect theory, individuals make choices that systematically deviate from how the idealized agent would behave. Such behavior is a product of the application of judgmental heuristics that produce cognitive errors. These errors are neither a product of poor incentives nor wishful thinking, and they cannot be overcome by learning. Actual choice behavior persistently deviates from the conventional norm established by subjective expected utility theory (Altman, 2010).

Under subjective expected utility theory, the utility function consists of a positive domain, where gains and losses are assumed equal with regard to utility. Individuals are assumed to estimate their utility in terms of states of wealth, subject to diminishing returns. Under prospect theory, individuals evaluate potential relative losses and gains in their decisions rather than an absolute level of wealth, a kink exists in the value (utility) function, and the slope of the value function is steeper for losses than for gains (Kahneman and Tversky, 1979). Reference points serve to frame the decision parameters. Thus, gains and losses are evaluated both separately in separate mental accounts (see below and refer to Thaler, 1985), and relatively, as opposed to simultaneously and in terms of absolute values or states of wealth, as under subjective expected utility theory (see Altman, 2010, and Thaler, 1985). Kahneman and Tversky (1979) model the value function in two segments according to:

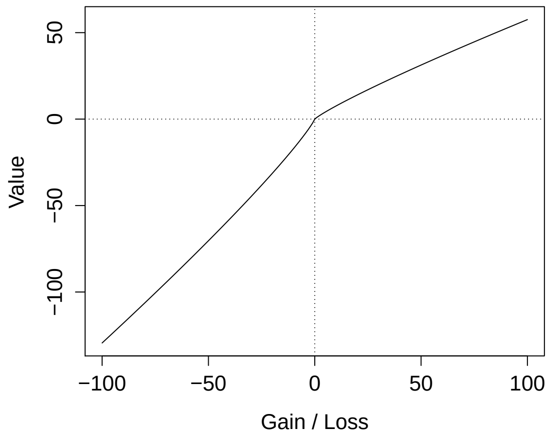
$$\nu(x) = \begin{cases} x^{\alpha} & \text{for } x \geq 0, \\ -\lambda(-x)^{\beta} & \text{for } x > 0. \end{cases} \tag{5.1}$$

where each prospect[16] x yields a value of $\nu(x)$, α and β are the shape parameters for the power utility function, and λ is the coefficient describing the difference in slopes of the positive and negative arms of the value function (see e.g. Hastie and Dawes, 2010). The power function describes constant-relative risk aversion (CRRA), see Prelec (2000) for details of differing shapes of utility functions under prospect theory.

[16]A prospect designates a state-contingent outcome. It is a course of action, the outcome of which depends on which state of nature is the true one. Formally, prospects map states to the reals, describing the resulting outcome for every state if that state is the true state (Wakker, 2010).

After empirical analysis, Tversky and Kahneman (1992) estimated parameters of $\alpha = \beta = 0.88$ and $\lambda = 2.25$; the alpha and the beta describe the degree of concavity or convexity of the positive and the negative arm of the value function, and the lambda describes how much the two differ. An example of the value function using these values is shown in Figure 5.1. Values are dependent on gains and losses, not on absolute values. The value function is concave in the domain of gains, convex in the domain of losses, and steeper for losses than for gains. The kink in utility in the origin of the graph in Figure 5.1 shows that people are usually risk averse when dealing with gains, but loss averse in the domain of losses. Individual investors are thus both risk-seekers and risk-averters at the same time. This is exhibited by their investing behavior, when individuals buy bonds, mutual funds, and insurance policies, acting in a risk-averse fashion. At the same time they buy individual stocks, options, and lotteries, and act as if they were risk-seeking (Yazdipour and Howard, 2010).

Figure 5.1: Prospect Theory Value Function



Kahneman and Tversky define two stages within prospect theory: an editing phase, which consists of a preliminary analysis of the offered prospects, and evaluation stage, in which the edited prospects are evaluated and the decision is made according to the highest value of a prospect (Kahneman and Tversky, 1979). According to prospect theory, this evaluation is made relative to some (potentially changing) reference point (in empirical analysis, usually either the market value of the asset at inception of the trade or the market value immediately preceding the current period are chosen), differing from expected utility theory, in which a rational agent is indifferent to the reference point. Recent studies also find confirmation of the formation of such varying reference points (Kliger and Kudryavtsev, 2008) amongst investors in corporate securities.

The utility function that underlies these valuations is more or less S-shaped around the reference point (see Figure 5.1) and reflects risk aversion when gains are made and risk seeking behavior when losses are experienced. In the gain domain, each additional unit of return is associated with an ever decreasing amount of utility, consistent with classical log-utility theory (Kahneman and Tversky also model the value function in the gain domain with a power function).

In the loss domain, prospect theory explains the disposition to sell winners and ride losers as investors framing their choices as a choice between two lotteries. The investor can a) either sell the stock now, realizing what would otherwise have been a paper loss, or b) hold the stock for one more period, given the investors subjective odds between losing an additional sum of money or breaking even. Since the choice between these two options is within the concave section of the value function, prospect theory predicts that choice b) will be selected, and that the investor will continue to hold a losing stock. Locke and Mann (2005) identify overconfident traders as those who, in a Bayesian sense, overweight the precision of their prior probability distributions of gains and losses, with new information continually entering into the decision to close the trade and thereby realize a gain or loss. These traders are thus the most likely to hold on to losing trades and exhibit a disposition effect. Coval and Shumway (2005) find that losing positions are held longer than gains among professional traders and interpret this as evidence for loss aversion.

The disposition effect as a development of prospect theory is the mainstream approach to the theoretical foundations underlying this particular behavioral bias. However, alternative approaches have been formulated, even if the body of theory as a counterpart to prospect theory is still small. Odean (1998a) introduces the theory that irrational beliefs in mean reversion could explain the disposition effect, as investors may believe that currently losing stocks will soon outperform today's winning stocks. If investors still hold to this belief despite contrary evidence, it could explain disposition effects. Hung and Yu (2006) base a further development of the Odean's idea as a formalized theory of the disposition effect.

Barberis and Xiong (2009) find that some aspects of prospect theory do not explain the disposition effect fully (in particular, they posit that investors derive utility only from realized gains or losses but not from paper gains or losses), and Kaustia (2010b) goes as far as to state that *"...prospect theory is unlikely to explain the disposition effect"*, but his view is definitely in the minority in the body of published analysis of the disposition effect, which generally accepts its heritage as stemming from prospect theory.

Recent research by Summers and Duxbury (2007) finds that merely experiencing gains or losses is not sufficient to produce a disposition effect. They support recent theoretical claims that prospect theory alone cannot explain the disposition effect. By taking the emotions of rejoicing or regret into consideration after having gained or lost during investing or trading, they argue that behavior induced by these emotions is sufficient to produce a disposition effect (Summers and Duxbury, 2007).

5.1.2 Mental Accounting

As described above, prospect theory can explain why investors are reluctant to hold profitable investments and sell unprofitable ones. Under prospect theory, gains and losses are evaluated both separately in separate mental accounts. The concept of mental accounting was proposed by Richard Thaler to provide a framework for how decision makers frame the gambles investors make. Thaler (1980), introduces the concept of mental accounting and continues in Thaler (1985), formulating the theory that decision makers separate their decisions into separate accounts, apply prospect theory rules to each account individually and ignore the fact that the accounts are actually interrelated. Under mental accounting considerations, investors are likely to refrain from readjusting their reference points for a stock, only establishing the reference point when the mental account is opened when the stock is purchased.

Empirical evidence shows that during the process of mental accounting, people engage in a concept described as ”narrow framing”, see e.g. Barberis and Huang (2001) in the current context, paying attention to gains and losses from individual stocks rather than over the total portfolio. Kahneman and Lovallo (1993) find that decision makers often consider the problem at hand in a unique fashion and isolate the choice that has to be made from other relevant issues or related choices. The consequences are either an overly cautious attitudes to risk (if the decision maker fails to consider risk-mitigating effects of aggregation of related risks within the current context) or overly optimistic forecasts. The latter result from an anchoring bias due to predictions being made under the influence of previous (not necessarily related) attitudes to problems (Kahneman and Lovello refer to this as the “inside view”). The two biases occur in a conflicting fashion, leading to cognitive difficulties regarding the decision at hand.

5.1.3 Regret Avoidance

The concept of narrow framing is taken by Barberis and Huang (2001) to imply that the concern for narrowly framed gains and losses may reflect regret over a decision to buy a stock that has performed badly as a source of utility not associated with consumption, i.e. investors get direct utility not only from consumption, but also from gains and losses in the value of individual stocks that they own. Regret is induced by closing a mental account associated with a loss-making stock, which is an emotion that people tend to avoid (Shefrin and Statman, 1985). Kahneman, Tversky and Thaler argue that the emotion of regret is stronger than the emotion of pride when closing a mental account related to a stock that has performed well. Those investors who show the disposition effect are assumed to be reluctant to realize both gains and losses.

5.1.4 Self-Control

This reluctance to realize losses is explained by Thaler and Shefrin (1981) as an intrapersonal conflict between an individual who is simultaneously a farsighted planner and a myopic doer (Thaler and Shefrin, 1981). The "doer" persona embodies the emotional reactions connected with regret and pride, and the "planner" provides the willpower to do the rational thing and avoid mental accounting or succumbing to the disposition effect. Thaler and Shefrin argue that the rational planner is often overpowered by the emotional doer, thus failing to act rationally. As a consequence, failing to exhibit self-control entices an investor to exhibit the disposition effect.

This concurs with the psychological concept of Two Systems of Reasoning (Sloman, 1996), or the Dual Process Theories (Stanovich and West, 2000). The two systems of reasoning according to Sloman (1996) are the associative system and the rule-based system. Stanovich and West (2000) call their two systems *System 1* and *System 2*. Common to the categorizations are that cognitive functions under the first System (or the associative system) include imagination, intuition and fantasy. System 2 involves deliberation, formal analysis and verification. Kahneman (2003) describes System 1 processes as *"...typically fast, automatic, effortless, associative, implicit ..., and often emotionally charged; they are also governed by habit and are therefore difficult to control or modify."* System 2 (or the rule-based system) in contrast is rule-based and therefore slower, but deliberately controlled. The psychological reasoning that leads to the disposition effect resulting from lack of self-control can be attributed to the fact that System 1 thinking often overrides System 2 decision-making due to limited mental resources (Kahneman, 2003). When input from System 1 is overruled by System 2, a form of cognitive dissonance takes place (Festinger, 1957) and makes rational decision more difficult or impossible - the self control that follows on from System 2 decision-making is inhibited by emotional decisions by System 1 and leads to the observed behavioral biases, such as the disposition effect in this case.

5.1.5 Current Research

Most published papers are able to prove the existence of the disposition effect on experimental markets as well as on for-real financial markets. Research evidence from funds has been published by Coval and Shumway (2005) documenting strong evidence of behavioral biases, the disposition effect in particular. Many of the papers published by authors such as Dhar and Zhu (2006) or Chou and Wang (2011) test whether trading frequency is related to returns or the disposition effect and whether increased trading prowess leads to lower disposition effects. However, recent research into the occurrence of the disposition effect amongst professional fund managers, (e.g. Singal and Xu, 2011), refutes the hypothesis that professionals are immune to disposition effects and that they can actually be measured, but that there is a distinction between better-performing and worse funds (measured by returns) that correlates with the disposition effect.

Research focusing on professional traders (in contrast to amateur or retail investors) provides evidence on the one hand by Garvey and Murphy (2004) or Frino et al. (2004) that even professional traders exhibit the disposition effect. On the other hand, recent fund research by O'Connell and Teo (2009) indicates that professional traders, in contrast to individual retail investors, might not seem as prone to the disposition effect as other authors assert. On futures markets, which are highly liquid and trading is performed frequently (similarly to the FX markets we analyze) as opposed to investment by funds, Choe and Eom (2009) find that trader sophistication and trading experience tend to reduce disposition effects. DeWeaver and Shannon (2010) argue that the problem individual investors may have regarding the disposition effect is that of "waning vigilance": Investors may pay less attention to new information when making decisions concerning loss-making positions than rational decision-making behavior would require.

Rational explanations for evidence of holding losers and selling winners include portfolio re-balancing, transaction costs or tax motivations (Frino et al., 2004). However, we can exclude all three of these reasons in our tests of the disposition effect on an experimental market due to the absence of transaction costs or taxation. Portfolio rebalancing as referred to by Frino et al. is also specific to professional fund managers and not relevant to individual investors on an experimental market.

On the foreign exchange market we analyze, tax motivations are not relevant as these are not taken into consideration by floor traders in a financial institution. Similarly, portfolio rebalancing is not performed in such a setting, as proprietary traders as well as corporate treasury sales employees do not hold client portfolios. Transaction costs may play a role, but are often taken to be out of bounds in a mental accounting sense and are not directly associated with a trader's individual profit and loss, but rather only on an aggregate level for the entire trading operation. It is highly unlikely that an individual trader would be interested in accounting for transaction costs unless directly financially motivated to do so.

In summary, we can discard the rational explanations of disposition effects on both the markets we study in the current chapter.

Similar to the costs associated with overconfident trading, the disposition effect as a behavioral bias is likely to be associated with substantial costs for companies employing such traders. This makes the topic of vital importance to risk managers acting in the interest of financial institutions in order to limit such avoidable costs. Obviously, with the evidence on hand since the paper by Shefrin and Statman (1985), the disposition effect is a measurable occurrence that costs investors and traders opportunity losses and missed gains doubly, on the upside and the downside. In previous research, Locke and Mann (2005) find that professional traders do exhibit the disposition effect, but they cannot find evidence of costs associated with this behavior - in contrast to Coval and Shumway (2005), who find evidence of costs associated with the disposition effect among professional traders.

Frino et al. (2004) also point out an important issue likely to impact any potential disposition effect of professional traders (particularly relevant to proprietary traders in financial institutions rather than fund managers): These types of traders are very often required to "go home flat" each day, i.e. to close out any outstanding long or short inventory positions they might have. This feature of proprietary traders is also observed by Kuserk and Locke (1993) and Manaster and Mann (1996). We expect to observe this behavior in our second study concerning FX traders, who are likely to conform to this behavior, even though not in the same form as pure futures traders who would not show any disposition effect on a time scale larger than intra-day. Even though the experimental market in the first study is set up in a similar fashion to a futures market, participants are not required to close out their positions, quite the opposite is the case as they act like investors over the complete investment horizon rather than floor traders.

We present our research on the disposition effect as a corollary to the overconfidence analysis performed in previous chapters on the described experimental market as well as on real foreign exchange (FX) data from a large multi-national financial institution. We find that the disposition effect is measurable on both the experimental and the real FX markets. Similarly to the findings regarding the inclusion of measurement of the overconfidence effect itself in risk management, we find no mention in literature of risk measurement techniques that include evaluations of the disposition effect. Therefore, we suggest a new addition to risk management mthodology, including evaluation of the disposition effect on an individualized or portfolio basis in risk management procedures in order to limit losses more effectively.

The objective of the current study is to analyze whether a) traders on the experimental soccer market and b) FX traders in a multinational financial institution exhibit loss aversion and subsequently to identify potential disposition effects. We present the data and analysis individually for the two independent studies and subsequently discuss common implications for risk management purposes.

Furthermore, we propose using the knowledge from the performed empirical tests to add points in the risk management measurement and limit process where disposition effects are measured, and to structure early warning systems that are targeted towards most likely candidates for intervention before losses are incurred.

5.2 Method

As mentioned in the Introduction, the de-facto standard methodology for tests of the disposition effect is Odean's (1998a) method derived from his analysis of 10,000 households with accounts at a large discount brokerage firm. We take Odean's methodology as outlined but do not need to make simplifying assumptions such as averaging values or only calculating on sale days. Rather, we implement the methodology in full to gain full advantage of the data available.

Odean (1998a) categorizes sales and purchases of individual stocks depending on whether the stock has risen or fallen. In particular, four measures of gains or losses are calculated: A realized gain (RG) is measured if a particular stock is sold for a gain since it was purchased. Likewise, a realized loss (RL) is counted if the stock in question is sold for a loss. The point of reference can be either the initial purchase time or the previous period. As our time horizon is only three weeks for the experiment, we use the initial purchase times as the reference points throughout against which gains or losses are measured.

For each time period (for Odean it is trading days, in our experiment it is the standardized time slice) paper gains (PG) or paper losses (PL) are accrued if the market value of the stock in inventory rises or falls from one time period to the next (excluding the final period, where gains or losses are realized).

A point in time can include both a realized gain/loss and a paper gain/loss if partial sales are made from inventory of a particular stock or if stock inventories are increased. In the case of a partial sale, the part that is retained receives a marker for either a paper gain or loss, and the sold portion is recorded as a realized gain or loss.

For each realized or paper gain/loss, a value of 1 or 0 is recorded, depending on the action, e.g. if a stock is bought and rises in value in the next period, it receives an PG of 1. If it rises again in the next period, another 1 is recorded for the PG variable. If it sinks in value, the PL variable records a 1 for this particular period. Analogously, when the stock is sold for a gain, the RG variable receives a 1, otherwise the RL variable is set to 1. The example in Table 5.1 from our data sample illustrates the methodology.

After the total number of realized and paper gains/losses are calculated for each participant, we compute the proportion of gains realized (PGR) and the proportion of losses realized (PLR), and from them the disposition effect (DE) as the difference between PGR and PLR as shown in Equations 5.2 and 5.3, following the notation in Kumar (2009).

$$DE_i = PGR_i - PLR_i, \quad \text{where} \tag{5.2}$$

$$PGR_i = \frac{N_{gr}^i}{N_{gr}^i + N_{gp}^i}, \quad PLR_i = \frac{N_{lr}^i}{N_{lr}^i + N_{lp}^i} \tag{5.3}$$

where DE_i is the calculated disposition effect, PGR_i and PLR_i are the proportions of realized gains or losses per stock put in relation to the total number of paper plus realized gains or losses, N_{gr}^i (N_{lr}^i) is the number of trades by investor i with a realized gain (loss), and N_{gp}^i (N_{lp}^i) is the number of potential or paper trades for investor i with a gain (loss).

The disposition effect DE per investor i occurs when $\mathrm{PGR}_i > \mathrm{PLR}_i$. A high PGR is indicative of losses that are sold too early, as the proportion of realized gains to total gains is high. The converse is true if the PLR is low - the numerator (the realized losses) are smaller in relation to the denominator (the total number of losses). In this case one would observe a disposition effect for investor i.

5.2.1 Study 1: Experimental Market

The following example illustrates the methodology as applied to our experimental market data:

Table 5.1: Disposition Effect: Soccer Example

User	Time Period	No. of Stocks	Stocks cum.	Price	PG	PL	RG	RL
008b01475d	5.10	100	100	6.25				
008b01475d	5.11		100	6.5	1	0		
008b01475d	5.12	−50	50	7	1	0		
008b01475d	5.13		50	5	0	1		
008b01475d	5.14	+20	70	4	0	1		
008b01475d	5.15		70	5	0	1		
008b01475d	5.16	−70	0	7	0	0	1	0
Sum					2	3	1	0

Note: This table shows the individual stock transactions in column *No. of Stocks* for an anonymized user identified by the hash key 008b01475d. The time slices are in decimal form on the 5th day beginning in decimal form at time 10. The column *Stocks cum.* shows the cumulative inventory of stocks at the respective time interval dependent upon the purchases and sales in the preceding column. In the last line no paper gain is recorded, as the stock inventory is completely sold and the gain is realized.

The percentage of gains realized (PGR) and losses realized (PLR) are then calculated from the results in the sum row of Table 5.1 and from these values the disposition effect (DE):

$$PGR_{008b01475d} = \frac{1}{1+2} = 0.33, \quad PLR_{008b01475d} = \frac{0}{0+3} = 0$$

$$DE_{008b01475d} = 0.33 - 0 = 0.33$$

In the example shown in Table 5.1, a DE value of 0.33 confirms the hypothesis of a positive disposition effect, the exemplified participant sells his gains more often than his losses.

We apply this methodology to each active participant and evaluate a single disposition effect value for each participant over all trades in all (up to 16) different stocks.

5.2.2 Study 2: FX Market

The general methodology that is applied to the FX market trades is the same as outlined above with one major exception being that disposition effects are calculated per portfolio instead of per trader. Several traders book their trades into the same portfolios which are each structured according to a certain style or purpose. The market values of each deal in each portfolio are compared per day with the initial market value for each deal. For each paper gain/loss of a deal (market value of the deal is above/below the initial market value of the deal during its lifetime) we assign a 1 in the paper gain/loss column and sum up per portfolio. The ending market value at the time of sale is compared to the initial market value to determine whether a realized gain or loss has taken place. We illustrate the methodology with a similarly structured example to the previous one:

Table 5.2: Disposition Effect: FX Example

Deal Number	Date	Portfolio	Market Value	PG	PL	RG	RL
11111	18.06.20xx	ABCD	$12,138,384.90$				
11111	19.06.20xx	ABCD	$12,057,924.20$	1	0		
11111	20.06.20xx	ABCD	$11,886,965.19$	0	1		
11111	21.06.20xx	ABCD	$12,920,306.08$			1	0
22222	18.06.20xx	ABCD	$300,706.35$				
22222	19.06.20xx	ABCD	$-18,090.45$	0	1		
22222	20.06.20xx	ABCD	$268,833.80$	0	1		
22222	21.06.20xx	ABCD	$269,985.83$			0	1
Sum				1	3	1	1

Note: This table shows the changing market values for two FX transactions that were initiated before 18.06.20xx with fictitious deal numbers 11111 and 22222 in portfolio ABCD. The deals are realized on 21.06.20xx, and the portfolio DE is calculated from the sum of the PGRs and PLRs of the individual deals within the portfolio.

The percentage of gains realized (PGR) and losses realized (PLR) are then calculated from the results in the Sum row of Table 5.2 and from these values the disposition effect (DE):

$$PGR_{ABCD} = \frac{1}{1+1} = 0.5, \quad PLR_{ABCD} = \frac{1}{3+1} = 0.25$$

$$DE_{ABCD} = 0.5 - 0.25 = 0.25$$

In the example shown in Table 5.2, a portfolio DE value of 0.25 confirms the hypothesis of a positive disposition effect, the exemplified traders who book into portfolio number ABCD sell their gains more often than their losses.

We apply this methodology to each deal in each portfolio, evaluating the individual market values each day and evaluate a single disposition effect value for each portfolio.

5.2.3 Participants and Data

Study 1: Experimental Market

As described in previous chapters, the 340 final participants on the experimental market form the final sample of participants for this analysis.

These 340 participants completed 48,757 trades in the three weeks of operation of the market. The number of trades per time period varied strongly (from 1 trade in 3 weeks to over 3,000 trades in the same time period), and participants were, as on a for-real futures market, able to enter or exit the market at any time. The compounding computational difficulty was to assign a market value to each participants portfolio at each point in time. Therefore, in order to accommodate these facts and to ensure necessary comparability at each point in time, the trade data was standardized. This was done by dividing each trading day into 100 equal time slices (i.e. each standardized time slice equals 14.4 minutes, resulting in 2702 time steps over the course of the market duration, from a decimal date of 3.49 (in decimal form, slightly before midday on the third of June) to decimal date 30.50 (exactly midday on June 30), after the last redemption trade was executed.

The value of 100 time slices per day was chosen to provide a reasonable time that would not be too short as to induce staircasing in the prices (when there are too few price changes in a certain time frame), and also not too many time slices to render the analyses too computationally laborious. This assured that a valuation of each portfolio component for each participant was calculated at each of the 2,702 time steps. If a participant traded a single stock more than once within a time slice, each transaction was valued at the correct transaction price and summed up within the time slice. This ensured that neither a FIFO or LIFO method was necessary (as other authors have found necessary under the same circumstances), but rather the actual purchase/sale price was used for valuation purposes.

Study 2: FX Market

The data set comprises 233,087 individual FX deals, transacted during a period of two years. Deals have maturities from one day to several months. A mark-to-market valuation is available for each deal on each day of its lifetime. After removing all trades with missing or incomplete data, 164,037 deals form the final sample. These deals are contained within 47 individual portfolios of differing sizes and business purposes. The number of deals per portfolio ranges from 2 to 41,817.

5.2.4 Research Design

Study 1: Experimental Market

As outlined in the general part of the Methods section, the following steps were performed:

1. The standardized price matrix P (2,702 time steps x 16 teams) was derived from the sum of all prices for each of the 16 stocks over their individual time frame
2. A vector of trades T_n was generated and standardized per participant n
3. A standardized cumulative vector of trade inventory I_n was generated per participant n
4. A standardized difference matrix Δ was calculated from the price matrix, the individual differences being calculated as the price at each time step minus the initial price P_t - P_0
5. Vectors of paper gains PG_n and paper losses PL_n were generated per participant n containing a 1 for each time the cumulative stock inventory multiplied by the

price delta $I_{nt} \times \Delta_t$ was greater/less than 0 and a 0 for each time $I_{nt} \times \Delta_t$ was less/greater than 0:

$$PG_{nt} = \begin{cases} 1 & I_{nt} \times \Delta_t \geq 0 \\ 0 & I_{nt} \times \Delta_t < 0 \end{cases}, \qquad PL_{nt} = \begin{cases} 0 & I_{nt} \times \Delta_t \geq 0 \\ 1 & I_{nt} \times \Delta_t < 0 \end{cases} \tag{5.4}$$

6. Vectors of realized gains RG and realized losses RL were generated containing a 1 for each time the stock transaction vector multiplied by the price delta $T_t \times \Delta_t$ was greater/less than 0 and a 0 for each time -$T_t \times \Delta_t$ was less/greater than 0[17]:

$$RG_{nt} = \begin{cases} 1 & -T_{nt} \times \Delta_t \geq 0 \\ 0 & -T_{nt} \times \Delta_t < 0 \end{cases}, \qquad RL_{nt} = \begin{cases} 0 & -T_{nt} \times \Delta_t \geq 0 \\ 1 & -T_{nt} \times \Delta_t < 0 \end{cases} \tag{5.5}$$

7. PGR_n and PLR_n were calculated per participant n according to Equation 5.3

8. Finally, DE_n was calculated per participant n according to Equation 5.2

9. Statistical significance of the results was determined with a t-test following Odean (1998a) with

$$t = \frac{PGR - PLR}{\sigma} \tag{5.6}$$

where

$$\sigma = \sqrt{\frac{PGR(1 - PLR)}{N_{rg} + N_{pg}} + \frac{PLR(1 - PLR)}{N_{rl} + N_{pl}}} \tag{5.7}$$

where N_{rg} denotes the number of realized gains and N_{pg} the number of paper gains, and N_{rl} the number of realized losses and N_{pl} the number of paper losses per participant n.

[17]The stock transaction vector T is multiplied by -1 to ensure that the sign of the realized gains or losses are correct, e.g. if stocks are sold, the transaction sign is negative, multiplied by a positive Δ this would yield a negative value, even though it is a realized gain. Thus the sign needs to be changed from negative to positive, so that a realized gain is positive and a realized loss is negative.

Study 2: FX Markets

After data cleaning, we used the mark-to-market values of each trade in each portfolio, sorted by date. The following steps were performed:

1. Vectors of sorted market values V_i per portfolio were generated

2. Vectors of daily differences in market values were generated for each deal:

$$\Delta_{ik} = V_k - V_0 \tag{5.8}$$

3. Similarly to study 1, vectors of paper gains PG_i and paper losses PL_i were generated per portfolio i containing a 1 for each time the market value at each time bucket t was greater than the initial market value of the deal or 0 for each time that the difference in market value was less than the initial market value of the deal:

$$PG_{ik} = \begin{cases} 1 & \Delta_{ik} \geq 0 \\ 0 & \Delta_{ik} < 0 \end{cases}, \qquad PL_{ik} = \begin{cases} 0 & \Delta_{ik} \geq 0 \\ 1 & \Delta_{ik} < 0 \end{cases} \tag{5.9}$$

4. Vectors of realized gains RG_i and realized losses RL_i were generated containing a 1 for each time a deal k was sold at a profit, or 0 when a deal was sold at a loss. A gain or loss was defined as the difference between the beginning and the ending market value of each deal:

$$\Delta_i = V_{ik} - V_{i0} \tag{5.10}$$

$$RG_{ik} = \begin{cases} 1 & \Delta_i \geq 0 \\ 0 & \Delta_i < 0 \end{cases}, \qquad RL_{ik} = \begin{cases} 0 & \Delta_i \geq 0 \\ 1 & \Delta_i < 0 \end{cases} \tag{5.11}$$

5. PGR_i and PLR_i were calculated per portfolio i according to Equation 5.3

6. Vectors of returns R_i for each deal in each portfolio were generated as the ending market value minus the beginning market value divided by the beginning market value:

$$R_{ik} = \frac{V_{ik} - V_{i0}}{V_{i0}} \tag{5.12}$$

7. Finally, DE_i was calculated per portfolio i according to Equation 5.2
8. Statistical significance of the results was determined with a t-test for proportions in the same fashion as in study 1 (Analysis of an experimental market, see Equations 5.6 and 5.7)

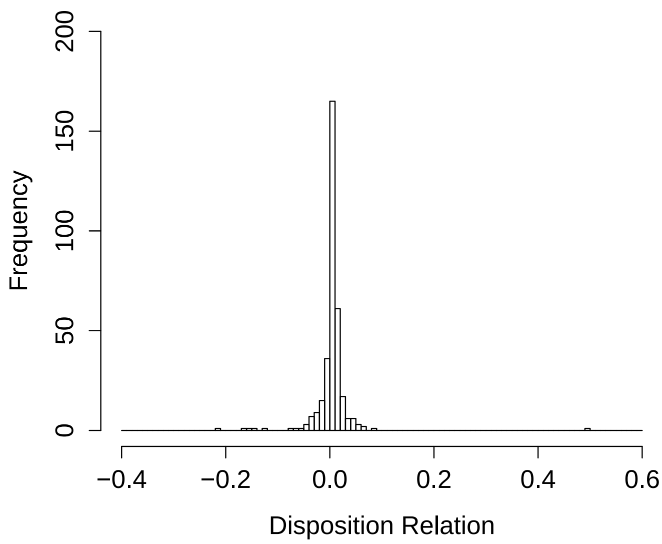
All calculations were performed using the open source statistics software R (R Development Core Team, 2011).

5.3 Results

5.3.1 Study 1: Experimental Market

The disposition effect was calculated as the proportion of gains realized minus the proportion of losses realized ($DE = PGR$ - PLR) for each of the 340 participants. On an individual level, 262 participants (77%) have positive disposition effects and 78 participants (23%) have negative values (none have exactly zero). The distribution of these values is shown in Figure 5.2. The distribution of disposition effects per participant has the following statistical characteristics: $\mu = 0.004$, $\sigma = 0.04$, skewness = 5.39 and kurtosis = 94.58.

Figure 5.2: Disposition Effect Frequencies

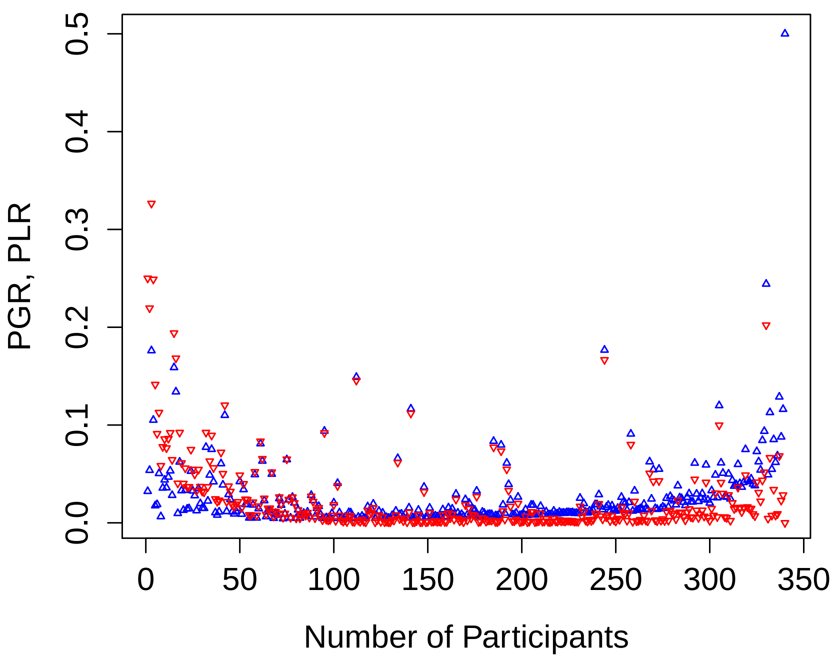


From the distributional characteristics as well as visually from Figure 5.2 it is obvious that the distribution of disposition effects has a mean close to zero and is right-skewed and extremely leptokurtic. There are many participants with small but positive disposition effect values. Figure 5.2 shows most of the distribution effect values bunched around the mean, slightly skewed to the positive side.

The t-test for proportions (test that the difference PGR_{total} - $PLR_{total} = 0$) does, however, confirm an overall disposition effect over all participants on an aggregate level. The overall disposition effect PGR_{total} - PLR_{total} = DE_{total} = 0.004, and the t-test confirms that this value is significantly different from zero, $t(339) = 2.19$, $p = 0.029$ (two-tailed), which is significant at the $p < 0.05$ level. This confirms the hypothesis of the existence of a disposition effect in aggregate over all participants.

Figure 5.3 shows the PGR and PLR values in blue and red, respectively, for each of the 340 participants. Many of the PGR and PLR values are close to each other in absolute value, which can also be determined from the tails of the distribution of individual DE values in Figure 5.2.

Figure 5.3: Relation of PGR to PLR per Participant



Note: The blue triangles pointing upward are the PGR value and the red triangles pointing downward are the PLR values per participant.

As the existence of disposition effects both in aggregate and on individual levels has been established, we test for potential connections between overconfidence as measured in previous chapters and the disposition effect. In order to find potential correlations, we regress the individual disposition effect values on each participant i's log-overconfidence indexes in the following form:

$$logOCI_i = \alpha + \beta DE_i + \epsilon_i \tag{5.13}$$

The results from the regression of the disposition effect values on each participant's log-OCI values are tabulated in Table 5.3.

Table 5.3: Regression Results of Disposition Effect on log-OCI

Coefficient	Estimate	Std. Error	t-Value	p-Value
(Intercept)	0.003	0.002	1.37	0.172
PRel (Disp.Eff.)	0.001	0.0005	1.92*	0.056
adj. R^2			0.007	
F-Stat			3.67*	0.056

Note: The table above shows the regression results from the regression of the disposition effect on the log-OCI in Equation 5.13. Ther results show a significant linear relationship between the disposition effect and overconfidence.
* denotes significance at the 10% level.

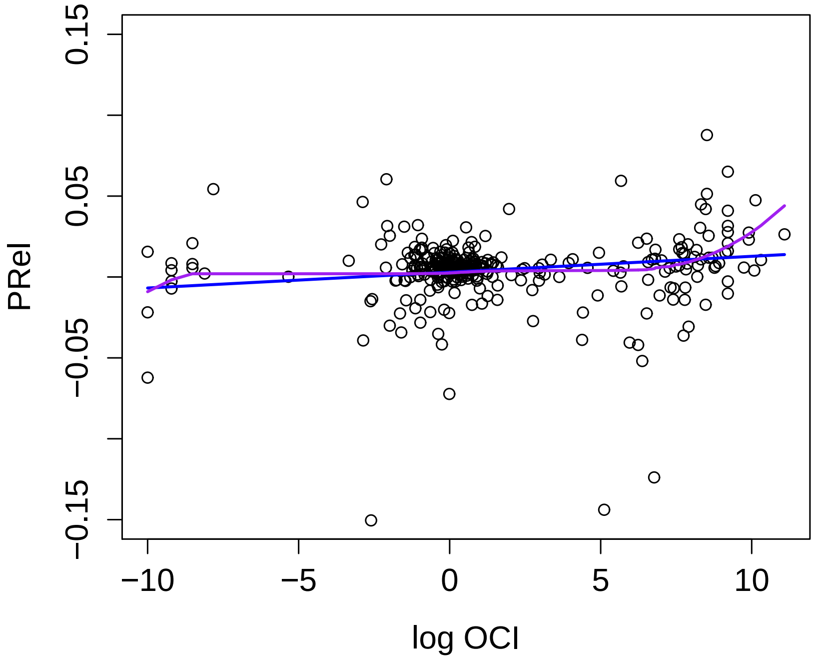
The regression results show that there is a linear relationship with p = 0.05, significant at the $p < 0.10$ level. From the visual inspection of the data displayed in Figure 5.4, it can be seen that the linear regression can be improved upon. In order to find curvilinear relationships that seem inherent in the data, we perform a cubic B-spline regression (using the function ns from the packages splines in R), see e.g. Hastie et al. (2009).

Visual inspection of the purple line in Figure 5.4 indicates a slightly better fit at both the negative and the positive extremes of the log-OCI values. In order to test whether the polynomial regression captures curvilinear relationships better than a linear regression, we perform the Davidson-MacKinnon J-test (see Davidson and MacKinnon, 1981)[18]

The results of the J-test are clear: Fitting the regressors of polynomial model into the linear model improves the p-value from p = 0.056 to p = 0.046, see Table 5.4 in the rightmost column. Vice-versa, fitting the linear regressors into the polynomial model does not improve fit (from p = 0.0135 to p = 0.01352). Thus, we can conclude at a

[18] The Davidson-MacKinnon J-test compares two models by including the fitted values of model 1 into model 2 and vice versa. If the second model contains the correct set of regressors compared to the first model, then including the fitted values of the second model into the regressors of the first model will provide a better fit than before. Fitting the regressors from the first model into the second model will then not improve the fit.

Figure 5.4: Regression of log-Overconfidence on the Disposition Effect



Note: The disposition effect on the y-axis is denoted PRel (Relation of Proportions), the blue line is the linear regression line, the purple curve the cubic spline regression with dof = 5.

Table 5.4: Davidson-MacKinnon J-test

	Disp.Eff. Estimate	Std. Error	t-Value	p-Value
M1	0.00	0.00	1.92*	0.056
M2	0.04	0.02	2.48**	0.014
M1 + fitted(M2)	1.00	0.49	2.00**	0.046
M2 + fitted(M1)	0.04	0.01	2.48**	0.014

Note: Model M1 is the linear regression of the log-OCI on the disposition effect and model M2 is the polynomial spline regression.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

$p < .05$ level that the polynomial regression provides a better fit, and therefore of the curvilinear relationship between overconfidence and the disposition effect. This shows that, in similar fashion to the kinked value curves in prospect theory, overconfidence tends to have a concave (in the negative domain) and convex (in the positive domain) relationship with the disposition effect. As an effect, when participants are very over- or underconfident, their tendency toward a disposition effect is marked. This result is of interest for those who are interested in detecting potential losses that are incurred by traders under the influence of the disposition effect.

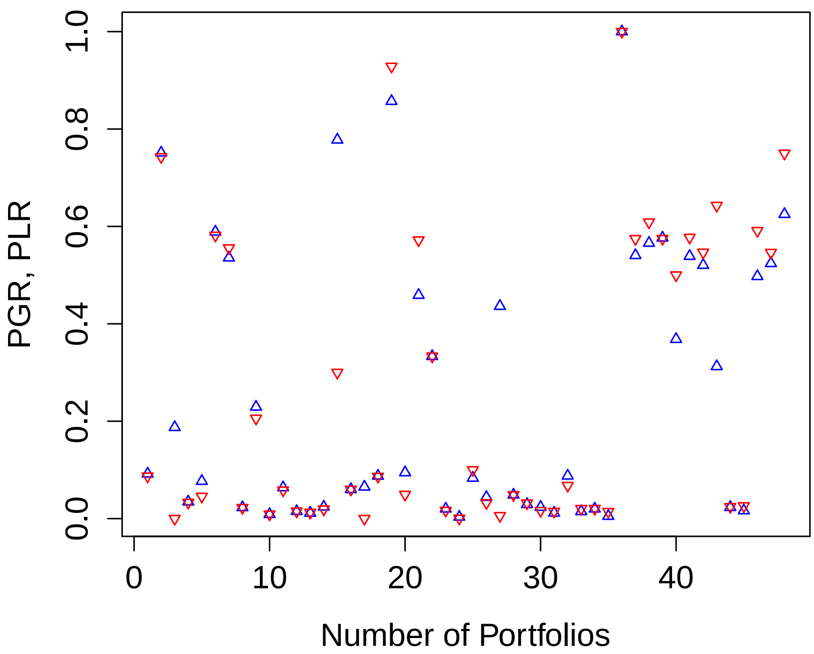
For risk management purposes, a focus should be put on monitoring those traders who prove to be very overconfident. The tendency toward exhibiting a disposition effect increases disproportionately when traders are very overconfident, thus incurring excess losses for the institution.

5.3.2 Study 2: FX Markets

In similar fashion to study 1, the disposition effect is calculated as the proportion of gains realized minus the proportion of losses realized ($DE = PGR$ - PLR) for each of the 47 portfolios. On a portfolio level, the situation is similar to at the aggregate level: 25 portfolios (53%) have positive disposition effects and 22 portfolios (47%) have negative values. Figure 5.5 shows the PGR and PLR values in blue and red, respectively, for each of the 47 portfolios. Even more than for the experimental market, most of the PGR and PLR values are close to each other in absolute value.

The distribution of disposition effects per portfolio has the following statistical characteristics: $\mu = 0.007$, $\sigma = 0.11$, skewness = 2.95 and kurtosis = 9.97. The FX DE distribution has a mean close to zero and is moderately right-skewed and leptokurtic. The average disposition effect in total is 0.007, but the t-test for proportions cannot reject the null hypothesis that this value is not significantly different from zero (t(46) = 0.45, p = 0.65, two-tailed).

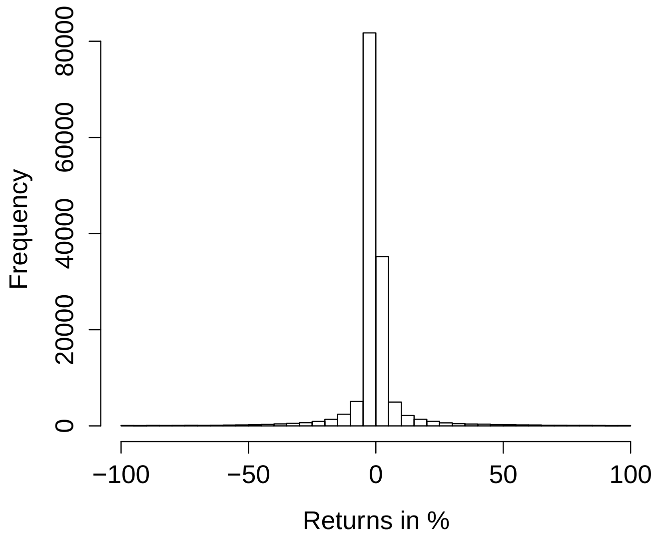
Figure 5.5: Relation of PGR to PLR per Portfolio



Note: The blue upward triangles are the PGR value and the red downward triangles are the PLR values per portfolio.

The returns of the 47 portfolios were calculated according to Equation 5.12. On average, the consolidated return over all portfolios was slightly positive at 0.34%, but 29 of the 47 portfolios (62%) had negative returns, as shown in Figure 5.6

Figure 5.6: Disposition Effect Frequencies



In order to determine whether there is a relationship between the disposition effect and the portfolio returns, we regress the returns on the disposition effect according to[19]:

$$R_i = \alpha + \beta_i DE_i + \epsilon_i \tag{5.14}$$

The results from the regression of the disposition effect values on each portfolio return values are tabulated in Table 5.3 for FX portfolios (study 2).

Table 5.5: Regression Results of Disposition Effect on Portfolio Returns

Coefficient	Estimate	Std. Error	t-Value	p-Value
(Intercept)	−0.14	0.26	−0.54	0.586
PRel (Disp.Eff.)	−4.75	2.31	−2.05**	0.046
adj. R^2			0.07	
F-Stat			4.22**	0.045

Note: The table above shows the regression results from the regression of the disposition effect on the portfolio returns in Equation 5.14. Ther results show a significant negative linear relationship between the disposition effect and returns.
** denotes significance at the 5% level.

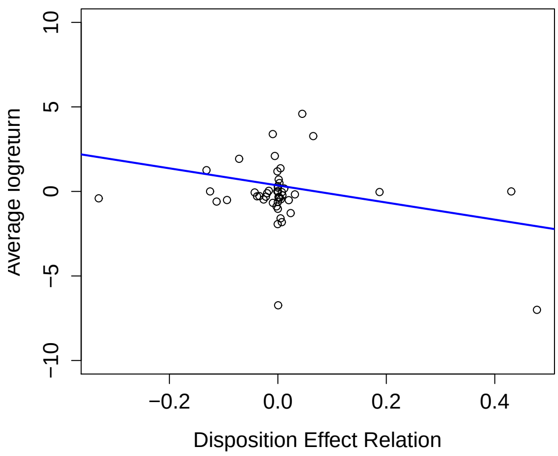
In contrast, in study 1 (the experimental market), the disposition effect values were

[19]The returns from the 47 portfolios (i) were winsorized at the 2% level to remove a very large outlier portfolio containing very few values from the regression with a huge positive return on just 2 deals.

regressed on the log-OCI in order to explore the relationship between overconfidence and the disposition effect. In study 2, portfolio returns were used in the regression, using Equation 5.14 rather than on the log-OCI according to Equation 5.13, as a corresponding overconfidence index was not calculated for the FX portfolios. However, the regression according to Equation 5.14 was also performed for study 1, and very similar results were achieved when regressing the disposition effect on the realized returns ($t(338)$ = -2.05, p-Value = 0.041, a negative relationship between realized returns and the disposition effect are significant at the 5% level).

The regression results for study 2 show a t-statistic of -2.05 and a p-value of 0.04, which is significant at the 5% level. The inverse relationship between the disposition effect and portfolio return can be seen in Figure 5.7. The more negative the portfolio return, the higher on average the disposition effect. This result holds for both data sets from study 1 and study 2 - subjects show a significantly higher disposition effect the lower the realized return on both the experimental market as well as the real FX market.

Figure 5.7: Portfolio Return vs. Disposition Effect



5.4 Discussion

The fact that the overall DE level on the experimental market is quite small at 0.002 may potentially be related to the fact that the experimental market analyzed was a closed market, i.e. the prices were all endogenously established (one participant's gain is another participant's loss) as a zero-sum game. Due to the fact that the aggregate disposition effect is endogenously determined by the holdings of all analyzed partipants, the disposition effect may in aggregate be systematically bounded.

Analyzing the individual *PGR* and *PLR* values (see Figure 5.3), we found that 77% of the participants have positive disposition effects and 23% have negative values. This provides substantial evidence of disposition effects at the individual level. Figure 5.3 shows graphically that the individual disposition values are very often close together which could be an indication of investment or trading styles. It seems that most participants were not prone to large disposition effects (borne out by the very slim tails in the *DE* distribution in Figure 5.2), but most held their gains slightly more than their loss-making positions. This gives rise to a positive disposition effect, but is interesting in that potentially, individual trading or investment styles due to psychological factors (such as those explored by questionnaire in Chapter 3, Overconfidence in Soccer) could be the more defining factor rather than the disposition effect. Interestingly, this effect is much more pronounced in the real-life FX portfolios. Figure 5.5 shows that the portfolio *PGR* and *PLR* values are extremely close together, which is an indication that the intentions and styles within portfolios are very similar, even though different traders book into common portfolios and inter-portfolio styles can be quite different.

The regression of the log-overconfidence values on the *DE* values is of particular interest. The nonlinear relationship between the disposition effect and log-overconfidence (the purple curve in Figure 5.4) shows that when overconfidence increases at a high level, the disposition effect increases more strongly. Likewise, when there is strong underconfidence present, the disposition effect decreases more strongly for a decrease in (over)confidence. It seems that in the middle of the distribution individual trading strategy is predominant, but towards the extremes in over- and underconfidence, the disposition effect increases or decreases disproportionately. In the sense that Locke and Mann (2005) describes, we find the disposition effect to be one good indicator of overconfidence bias in trading or investment situations.

The distributional characteristics of the disposition effects on the analyzed FX market deals are slightly different from the characteristics of the analyzed experimental market. The disposition effect is evenly distributed, 53% of the portfolios showing a positive effect. Also, the disposition effect in aggregate was not significantly different from zero. This could result from several effects: Several traders booked their trades into common portfolios (as is usual in such situations). Some traders could be more susceptible to the disposition effect than others, so a diversification of behavioral biases may reduce the disposition effect in such portfolios. Secondly, the different portfolios are set up with different trading intentions in mind. Those portfolios that contain held-to-maturity positions will necessarily show different dispositional characteristics than available-for-trading positions (in an IFRS accounting sense). Portfolios containing corporate sales positions will also have different characteristics.

As mentioned in the Introduction, several other authors such as Garvey and Murphy (2004) or Frino et al. (2004) have found the disposition effect amongst professional traders, but to be less pronounced than amongst amateurs (O'Connell and Teo, 2009). Kumar and Lim (2008) discuss narrow framing and its influence of investment decisions, and suggest that separate decisions lead to more narrow framing of decisions than simultaneous decisions[20]. They propose that clustering of trades, such as is likely to occur in trading situations such as the FX trading markets we analyze, lead to weaker disposition effects than e.g. amateur investors who do not trade so frequently over a long period of time. Our results concur with their findings and offer the explanation that narrower framing by amateur traders could lead to higher disposition effects than for professional traders.

We are in the unique position of being able to directly contrast amateur traders on an experimental market and professional foreign exchange traders acting on real-life markets. As other authors have found in recent studies, outlined in the Introduction section, we do find that in our samples professional traders are on aggregate less likely to exhibit such a pronounced disposition effect as the amateurs or retail investors, but we still find that portfolio returns are significantly connected with the disposition effect in professional trading. We find that those traders who are prone to the disposition effect definitely reduce their returns.

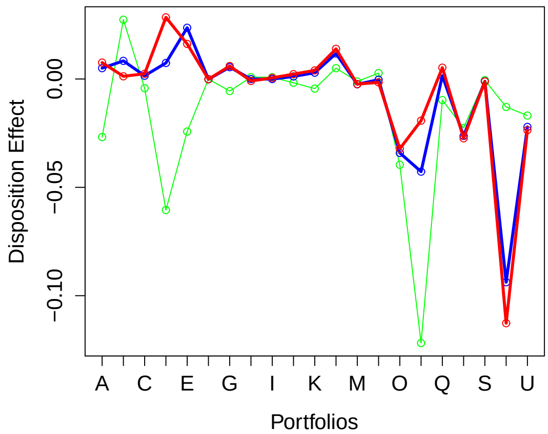
5.5 Robustness Checks

In order to verify that the above results were not achieved entirely due to chance, we analyze the robustness of the results, following the sample partitioning method described e.g. by Bishop (2006) or Witten et al. (2011). For the FX trades, first we winsorize the sample of porfolios at the 1% level, in order to use only portfolios with a reasonable number of deals, otherwise splitting each portfolio would skew results due to the small sample bias in portfolios containing very few deals. Of the total number of portfolios (47), we discard the the lowest 1% (28 portfolios containing 1,699 deals) by number of deals. 22 portfolios remain, containing 162,338 deals in sum.

Next, we partition each portfolio of the remaining 22 portfolios into the first 75%

[20]Narrow framing in an investment context implies that actors evaluate new risks in isolation rather than in the context of their entire portfolio, see e.g. Barberis and Huang (2001).

Figure 5.8: DE Total, 75% Partition and 25% Partition



Note: Blue line: Total portfolio disposition effect values. Red line: 75% partition values. Green line: 25% partition values. It can be observed that the 75% partition closely follows the 100% total, and that the 25% partition suffers from data problems due to the small sample size, despite having discarded the 28 smallest portfolios from the sample, leaving 22 portfolios containing 99% of all trades.

of its trades and the last 25% (by trade date) and recalculate each portfolio following the steps outlined in the Research Design section. Figure 5.8 shows the results of the disposition effect calculation using 100% of the trades in each portfolio (DE Total in Figure 5.8, as well as the disposition effect calculated for the first 75% and the last 25% of each portfolio the (75% partition and the 25% partition).

It is visually obvious that the 75% partitioned values are very similar to the total portfolio values, but the sample size for the 25% sample is too small to deliver accurate values. In order to ascertain a good fit for similarity, we regress the PGR and PLR results of the 75% partition and the 25% partition on the total portfolio PGR and PLR values, as well as the two partition values on each other, this creates twice the number of observations by using PGR and PLR values individually rather than only the DE value, and is less susceptible to outliers influencing the direction of the individual disposition effects:

$$PGR_{total} = \alpha + \beta PGR_{75\%} + \epsilon \tag{5.15}$$

$$PLR_{total} = \alpha + \beta PLR_{75\%} + \epsilon \tag{5.16}$$

$$PGR_{total} = \alpha + \beta PGR_{25\%} + \epsilon \tag{5.17}$$

$$PLR_{total} = \alpha + \beta PLR_{25\%} + \epsilon \tag{5.18}$$

$$PGR_{75\%} = \alpha + \beta PGR_{25\%} + \epsilon \tag{5.19}$$

$$PLR_{75\%} = \alpha + \beta PLR_{25\%} + \epsilon \tag{5.20}$$

The results of the regressions are tabulated in Table 5.6 and shown visually in Figure 5.9.

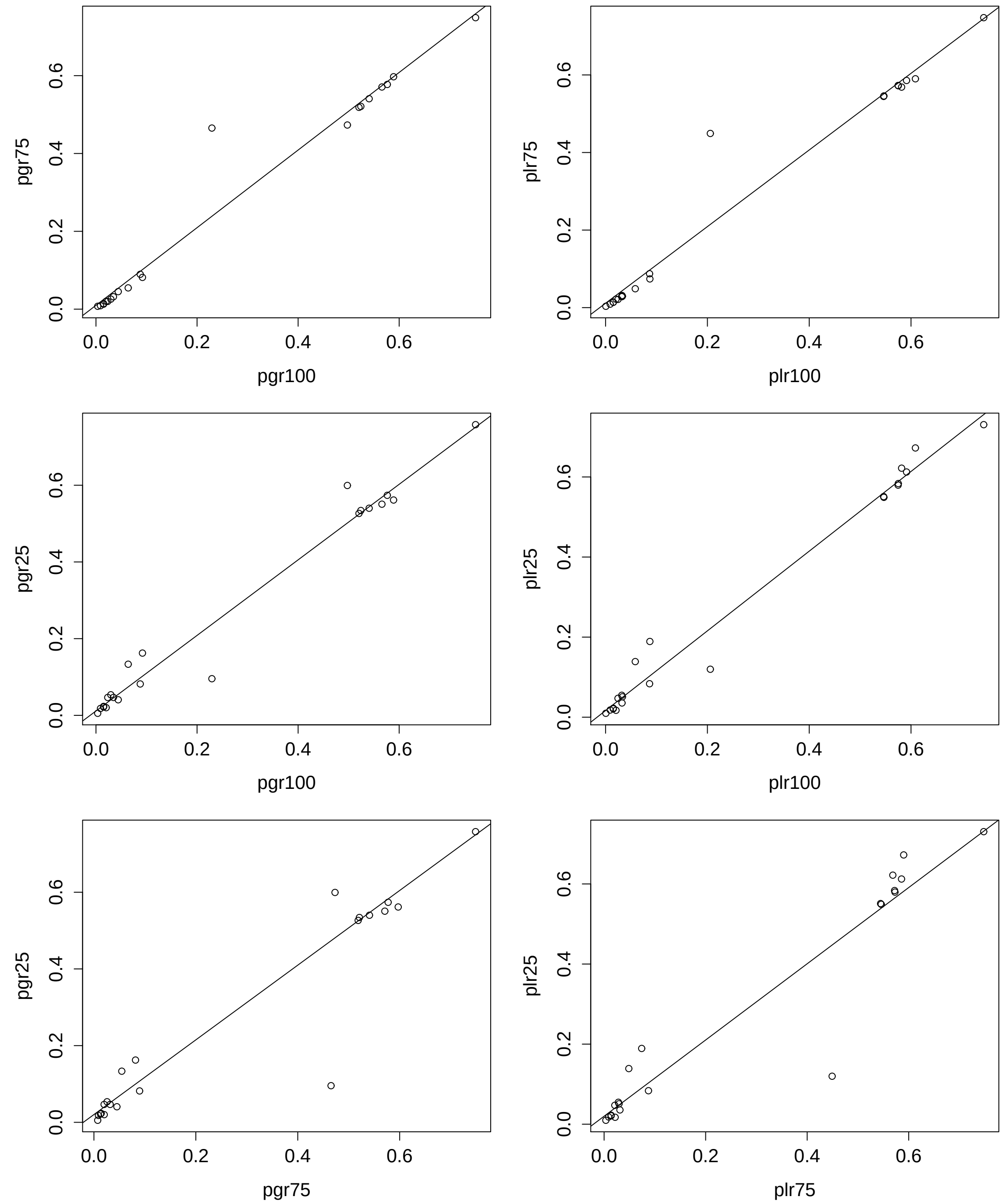
Table 5.6: Regression of Partitions on Total DE

Category	Estimate	Std. Error	t-Value	p-Value
PGR 75% on 100%	0.99	0.04	22.15***	< 0.001
PLR 75% on 100%	0.98	0.04	22.18***	< 0.001
PGR 25% on 100%	0.98	0.03	25.86***	< 0.001
PLR 25% on 100%	0.99	0.03	32.64***	< 0.001
PGR 25% on 75%	0.92	0.07	12.18***	< 0.001
PLR 25% on 75%	0.95	0.06	14.00***	< 0.001

*** denotes significance at the 1% level

The results of the regressions show that the *PGR* and *PLR* values for the 75% and 25% partitions are closely related to the total disposition effect and each other (see the individual graphs in Figure 5.9). The regressions of both partitioned subsamples on the total sample as well as on each other are all significant at the $p < 0.001$ level. Thus the partitions explain the full portfolio extremely well and provide a good robustness check of the disposition effect data, so we can be fairly confident that the disposition effect results were not achieved by chance alone but offer good explanations of the observed data.

Figure 5.9: PGR and PLR Regressions for 100%, 75% and 25%



Note: The total portfolios were split into two portfolios, the first portfolio containing the first 75% of all deals, the second portfolio containing the last 25% of all deals. The two top figures show the regressions of the disposition effects of the 75% portfolios on the total portfolios, the middle figures show the regressions of the 25% portfolios on the total portfolios, and the bottom figures show the regressions of the 25% portfolios on the 75% portfolios.

Risk Management

As outlined in the Introduction, Behavioral Finance has, to the best of the author's knowledge, not found its way into risk management practices or processes yet. As has been identified by other authors (see e.g. Hung and Yu, 2006), most studies performed on the disposition effect are empirical in nature and are concerned with the ex-post identification of disposition effects. The identification of behavioral biases is undoubtedly an important part of the functioning of financial markets, but as documented in the Introduction, most authors are of the opinion that susceptibility to behavioral biases such as the disposition effect cost money. In particular, this has been proven to be the case in developing markets (Chen et al., 2007).

Therefore, it seems logical to find ways to utilize the knowledge of how to identify potential sources of loss-making behavior before the fact and to set up early-warning mechanisms to combat the adverse effects. Even though authors such as Dhar and Zhu (2006) find that investor sophistication mitigates the disposition effect (that the disposition effect is weaker on aggregate amongst professional traders with high trading frequency is also borne out by our current analysis) we show that the aggregate view is not necessarily the optimum vantage point for this purpose.

On the individual level, our analysis shows that there are recognizable styles of trading amongst traders booking into different portfolios. Certain portfolios have been shown to systematically be prone to disposition effects (confirmed by the robustness checks that were conducted and outlined above). We propose that the risk management departments of financial institutions, where trading activities form a non-negligible part of profits or losses, should implement ongoing calculations of disposition effects as we have described in the section Research Design.

Early warning systems should be put in place on the basis of the results of the (maybe daily) evaluations of both variables PGR and PLR, as even the difference of these two variables is more imprecise than measurement of the input factors.

Additionally, we propose that management should be presented with ongoing disposition effect ratios either per trader or per portfolio, focusing on the top decile of most disposition-effect prone traders or portfolios as management information.

Lastly, more customized individual trader-level or portfolio-level stop-loss limits should be put into place on the most affected portfolios in order to limit potential losses before they start becoming a problem.

Even though Wong et al. (2006) do not find that personal characteristics determine the disposition effect directly, our previous research in earlier chapters show that overconfidence is shaped and is forecastable by personal characteristics or traits. The disposition effect as either a part of overconfidence or at least as a corollary to overconfidence should however be observed and managed in the same way as we have shown in this paper and previously that overconfidence should be managed in a risk management context. Therefore the personal traits that shape overconfident behavior are in our opinion also determining factors in the recognition and forecastability - and therefore in the early warning of - disposition effects.

5.6 Conclusion

In this paper we analyzed the disposition effect in two studies, one on an experimental market with non-professional retail investors, the other on a real-life high-frequency market with professional traders. Following the methodology initiated by Odean (1998a), we analyzed the 48,757 individual trades of 340 retail investors on the experimental market, and 233,087 foreign exchange trades contained within 47 portfolios from a multinational financial institution.

We found the disposition effect to be observable in aggregate for the market as a whole on both markets, but more pronounced in the experimental market with non-professional investors. We also found the disposition effect to correlate with overconfidence and with lower returns. Robustness checks substantiated our results and, based on these findings, we make recommendations for risk management to set in place measurement systems to identify personnel and aggregated portfolios for disposition effects and further to implement individualized limit systems to prevent the behavioral biases of individual traders or investors from losing too much of other people's money.

Further research should be performed in order to narrow down points of attack for identifying the most promising methods of correcly identifying disposition effects on an individual basis, integrating these identification methods into risk management practice, and to find out which methods prove to be the most efficient in limiting losses from the disposition effect.

Chapter 6

Overconfidence in Engineering

6.1 Introduction

In the previous chapters we have analyzed behavioral biases of traders in an experimental market and in a real market. Due to traders' common reputation transported in the media, one might assume that traders are inherently prone to overconfidence, so in order to contrast traders' behavioral biases with another group of participants, we research whether overconfidence expressed as miscalibration (see Lichtenstein et al., 1982) or expressed as the better-than-average effect (see Svenson, 1981) can be identified among professional engineers and engineering students.

Addionally, we research the perceptions of engineering professionals concerning their risk management self-assessments and abilities.

6.2 Method

Two questionnaires were given to same group of engineering students two months apart with the intention of identifying potential overconfidence and better-than-average effects based on an upcoming engineering project that was to be performed as an individual task by each of the students in the respective classes.

One questionnaire was provided to engineering professionals that aimed at identifying behavioral biases and risk management perceptions and attitudes.

6.2.1 Participants

Engineering Students Questionnaire 1

There were 118 respondents to questionnaire 1, their ages ranged between 17 and 49 years (M = 22.7 years, SD = 6.64 years). The subjects were in 5 different courses ((12. grade, 13. grade, 2. Semester Kolleg - ages about 2-21, 4. Semester Kolleg - ages about 22-23, and 4. Semester Werkmeister - adult, ages above 23), 96 (81.4%) subjects were male and 22 (18.6%) were female.

Engineering Students Questionnaire 2

The second sample changed slightly in composition, as one of the classes (4. semester Werkmeister) was not available for the retest. From the remaining sample that was comparable to those respondents to questionnaire 1, there were 60 respondents to questionnaire 2, their ages ranged between 17 and 35 years (M = 20.32 years, SD = 4.38 years). The subjects were from the same sample of courses (12. grade, 13. grade, 2. semester Kolleg, 4. semester Kolleg), 44 (73.3%) subjects were male and 16 (26.7%) were female.

Professional Engineers

There were 13 respondents to the questionnaire, their ages ranged between 34 and 63 years (M = 50.54 years, SD = 9.07 years). Subjects' work experience ranged between 9 and 41 years (M = 26.5 years, SD = 8.77 years). All subjects were male.

6.2.2 Research Design

Engineering Students

Two questionnaires were provided to the same grouping of engineering students (subsequently "questionnaire 1" and "questionnaire 2"); the second questionnaire was administered 2 months after the first one when the students had completed their individual engineering projects.

Engineering Students Questionnaire 1

The first questionnaire contained 28 items, of which 4 questions were open questions, 1 was dichotomous (male or female), and the remaining items were structured on a 5-part Likert-type scale (Likert, 1931). There were 2 parts to the questionnaire, a general part which posed questions containing items such as self-declared optimism levels, whether the students saw themselves as more academically successful, careful, or diligent than their class colleagues and whether they would prove be more professionally successful compared to their class colleagues and a specific part containing questions regarding students' perceived abilities concerning the upcoming engineering project. These questions were also specifically set to ask how students compared their own perceived abilities versus those of their class colleagues as their peer group:

General Part scales:

1. Demographic data: 3 items
2. Perceived general abilities (individual & related to estimations of peer group): 12 items, and
3. Perceived engineering abilities (individual & peer-group-related): 2 items

Specific Part scales:

1. Items concerning specific project estimates (time, effort, complexity): 7 items and
2. Individual and peer-group-related questions: 4 items.

Engineering Students Questionnaire 2

The second questionnaire, which was shorter in length, was administered 8 weeks after the first questionnaire, and structured to re-test some of the overconfidence-related items from questionnaire 1 and to test for data after the students' engineering project papers had been handed in. The questionnaire was also structured in a general part and a specific part, containing 3 demographic questions, 4 general items re-testing items from the general part of questionnaire 1, and 11 items concerning the recently completed effort invested in the engineering project and expectations concerning grading of the project.

Engineering Professionals

The questionnaire contained 40 items, structured in a general and a specific part of which 9 questions were open questions, 1 was dichotomous (male or female), and the remaining items were structured on a 5-part Likert-type scale. 3 items were demographic questions, 2 items concerned general attitudes to optimism and risk-taking, 19 items were related to individual self-assessment of abilities and behavior, and the rest (16 items) concerned own abilities in direct comparison with colleagues and peer groups.

Analysis

Factor Analysis was performed to identify common factors within questionnaires that could be extracted and overconfidence tested on the resulting factors. The factor analyses were performed using the standard functionality within the statistical software R (R Development Core Team, 2011).

6.3 Results

6.3.1 Students Questionnaire 1

Male vs. Female Engineers

An analysis of whether the sex of the subjects made a difference in their answers was performed via a logistic regression (see e.g. Borooah, 2002) of the form:

$$D = \alpha + \sum_{k=1}^{K} \beta_k Q_k + \epsilon, \quad k \in \{4,...,10\} \tag{6.1}$$

where the question 3 (male or female) was regressed on the following questions:

Q4: Do you tend to be an optimistic or a pessimistic person?

Q5: Do you achieve better or worse grades than your class colleagues?

Q6: Do you achieve better or worse adacademic results (regardless of grades) than your class colleagues?

Q7: Are you more or less content with your achieved academic results?

Q8: Are you more or less diligent in your project work than your class colleagues?

Q9: Are you slower or faster than than your class colleagues in project work?

Q10: Do you expect to earn more or less than your class colleagues later in professional life?

The results from the logistic regression are shown in Table 6.1.

Table 6.1: Logistic Regression Results

Question	Coefficient	Std. Error	z-value	p-Value
Q4: Optimism	0.42	0.32	1.29	0.196
Q5: Better grades	−0.40	0.45	−0.88	0.375
Q6: Better academic results	−0.44	0.43	−1.01	0.311
Q7: Content with acad. results	0.57	0.29	1.95*	0.051
Q8: Diligence	−0.12	0.30	−0.41	0.680
Q9: Slower/faster	−0.12	0.30	−0.39	0.690
Q10: Expect to earn more	0.91	0.37	2.40**	0.016

*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

The male students are neither generally more optimistic than their female counterparts, nor do they expect to receive better grades or achieve better academic results - the coefficients for these two values are negative, but not significantly so. However, they seem to be more content with the results that they actually achieve than female students. Men neither consider themselves to be more or less diligent than women or slower or faster in achieving project results, but the male students do show a form of (over)confidence insofar as each of them expects to earn more than the other students when professionally employed.

The direction of the sign of the coefficients does conform with traditional stereotypes, even though only two of the values are significant: optimism, contentedness with achieved results and expectations of future better earnings all tend to be attributes of the male students, whereas achieving better grades, better academic results and diligence tend to be more female attributes in our sample of engineering students.

Factor Analysis

The factor analysis was performed using the varimax orthogonal rotation methodology. A scree plot was generated in order to identify the correct number of factors to extract.

The scree plot shown in Figure 6.1 identifies a maximum number of 8 factors (PA1 to PA8) to be extracted (the principal factor extraction method from the statistical package psych in R).

Figure 6.1: Screeplot of Engineering Students 1

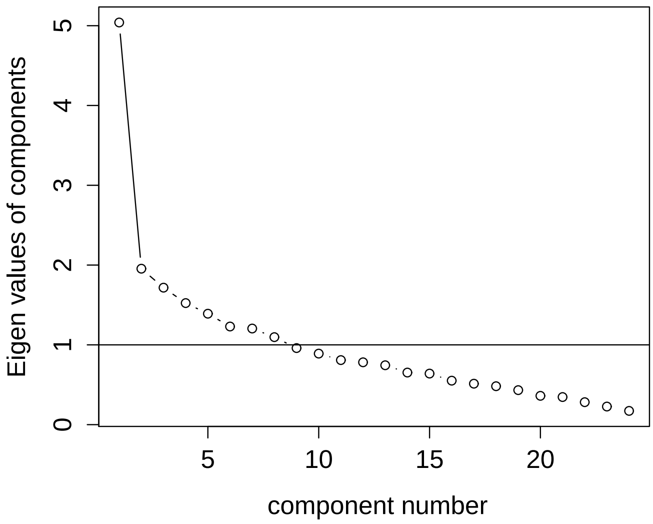


Table 6.2 shows the factor loadings on each of the 8 extracted factors. The interesting loadings are highlighted in boldface.

The factor analysis shows a high reliability - the Tucker Lewis Index of factoring reliability = 0.96, and the RMSEA index = 0.043. The fit based upon off-diagonal values is .98, and score adequacy is high, with correlation of scores with factors ranging from 0.77 (PA4 and PA6) to 0.93 (PA1). The multiple R^2 of the scores with factors ranges from 0.59 (PA6) to 0.87 (PA1)[21].

[21] The difference between principal factors and principal components is that in principal components analysis it is assumed that all variability in an item should be used in the analysis, while in principal factors analysis only the variability in an item that it has in common with the other items is used. In most cases, these two methods usually yield very similar results. However, principal components analysis is often preferred as a method for data reduction, while principal factors analysis is often preferred when the goal of the analysis is to detect structure (see e.g. Kim and Mueller, 1978a,b)

Table 6.2: Factor Analysis Results Engineering Students

Question	PA1	PA2	PA8	PA6	PA3	PA4	PA7	PA5
Q.4 Optimism	0.01	0.00	0.09	0.02	**0.47**	0.21	0.09	−0.10
Q.5 Better/worse grades	**0.88**	0.01	0.26	0.12	−0.02	0.07	0.06	0.08
Q.6 Better/worse academic results	**0.64**	0.07	0.31	0.24	0.10	0.10	0.17	0.02
Q.7 Content with acad. achievement	**0.56**	−0.05	0.04	0.17	0.06	0.08	−0.04	0.24
Q.8 More/less diligent	0.18	0.06	0.07	**0.62**	0.04	0.06	−0.01	−0.07
Q.9 Faster/slower project work	0.09	0.01	**0.67**	0.15	0.11	−0.04	−0.12	0.19
Q.10 Will earn more/less	0.26	0.06	**0.53**	0.00	0.10	0.12	0.10	−0.12
Q.11.1 Better/worse technical ability	0.47	−0.06	**0.49**	0.03	0.25	−0.12	−0.01	−0.01
Q.11.2 Better/worse time management	0.29	0.1	0.39	0.32	−0.06	0.04	−0.21	−0.01
Q.11.3 Better/worse foreign languages	0.02	−0.06	0.04	0.13	−0.08	**0.61**	−0.02	0.01
Q.11.4 Better/worse gen. education	0.11	0.04	0.11	0.02	0.25	**0.55**	0.12	0.02
Q.11.5 Better/worse musicality	0.03	0.02	−0.09	−0.01	0.15	0.32	−0.23	−0.14
Q.12 Better/worse ideal of engineer	0.11	−0.04	**0.55**	0.14	0.03	0.13	**0.27**	0.21
Q.13 Better/worse project result	0.03	−0.06	0.24	**0.62**	0.34	−0.03	**0.25**	0.18
Q.14 Expected grade for project	−0.18	0.15	−0.19	−0.64	−0.15	−0.16	0.03	−0.08
Q.15 Expected average grade	0.00	**1.36**	0.08	−0.07	0.00	−0.03	0.10	0.00
Q.16 Better/worse project than previous	0.02	−0.01	−0.04	0.26	**0.47**	−0.09	0.03	0.24
Q.17 Class colleagues will be better/worse	−0.16	−0.12	−0.37	−0.31	0.15	0.01	−0.07	0.12
Q.19 Forecast difficulty of project	−0.06	−0.11	−0.06	−0.03	−0.13	0.03	**−0.60**	0.18
Q.20 Necessary techn. ability	0.35	0.00	0.22	0.19	**0.48**	0.12	0.07	0.10
Q.22 Is project achievable	0.21	0.00	0.10	0.01	0.06	−0.05	−0.11	**0.62**
Q.23 Importance of project	0.06	−0.15	−0.06	0.26	0.15	0.12	0.25	0.14

Note: The Table above shows the results of a principal factors analysis with varimax orthogonal factor rotation. The values in the PA1 - PA8 columns are the principal axis factors, in order of their contribution to overall explained variance. The factor loadings shown in boldface are those that are extracted to areas of interest (commonalities).

With all items except the open questions, Cronbach‘s $\alpha = 0.12$. When items 2, 11.2, 15, and 17 (i.e. those items that were not significant in the factor extraction) are removed, Cronbach's $\alpha = 0.68$, a distinct improvement to a value that indicates reasonable but not excellent internal reliability of the questionnaire items[22].

The following factors could be identified:

PA1: Positive loading: Current academic ability in engineering

PA8: Positive loading: Self-confidence as an engineer and of future earnings

PA5: Positive loading: Good project result through diligence, will receive good grade

PA6: Positive loading: Diligence will result in good project work

PA3: Positive loading: Optimism in general abilities

PA4: Positive loading: General education

The factors that were extracted show that the items regarding self-perceived abilities and technical knowledge correlate highly (PA1), as well as self-perception regarding the level of education - a potential source of overconfidence. Another potential indication for overconfidence is in factor PA7: The better students think they are academically than their peer group, the less difficulty is associated with the project.

Overconfidence Analysis

On the basis of the factor loading indications of overconfidence, we test each of the relevant items against their means with a standard t-test for means.

[22] Classical test theory distinguishes three sources of variance: (a) true score variance, (b) error variance or measurement error, and (c) total scale variance, which is the sum of true score and error variance. Reliability is defined as the proportion of variance in an observed test score that is related to the true scores (McDonald, 1999). Cronbach's alpha (or coefficient alpha), see Cronbach (1951), is one of the most popular reliability measures of internal consistency of tests (Miller, 1995).George and Mallery (2003) provide the following rules of thumb for the acceptability of Cronbach's α : $> .9$ – Excellent, $> .8$ – Good, $> .7$ – Acceptable, $> .6$ – Questionable, $> .5$ – Poor, and $< .5$ – Unacceptable.

Table 6.3: Overconfidence Analysis Results

No.	Question	Mean	SD	p-Value	OC/WC
5	Do you normally achieve better or worse grades than your class colleagues?	3.26***	0.91	$p < 0.01$	OC
6	Are you of the opinion that your academic achievements are better or worse than those of your class colleagues (regardless of whether you achieve better grades or not)?	3.32***	0.84	$p < 0.01$	OC
8	Are you more or less careful when doing class project work than your class colleagues?	3.46***	0.88	$p < 0.01$	OC
9	In comparison with your class colleagues, are you normally faster or slower when doing class project work?	3.15*	0.87	$p = 0.06$	OC
10	Do you think you will earn more or less money than your class colleagues in the future?	3.47***	0.80	$p < 0.01$	OC
11.1	Do you have better/worse technical abilities than your class colleagues?	3.50***	0.84	$p < 0.01$	OC
11.2	Are you better/worse at time management than your class colleagues?	3.30***	0.93	$p < 0.01$	OC
11.3	Are you better/worse at foreign languages than your class colleagues?	3.15	1.11	$p = 0.14$	WC
11.4	Are you better/worsein general education than your class colleagues?	3.58***	0.78	$p < 0.01$	OC
11.5	Are you better/worse in musicality than your class colleagues?	2.83	1.19	$p = 0.12$	WC
12	I fulfill the professional role of an ideal engineer better/worse than my class colleagues	3.25***	0.74	$p < 0.01$	OC
13	Compared with your class colleagues, how well will you do on your class project? My project result will be better/worse than that of my class colleages	3.56***	0.69	$p < 0.01$	OC
14	Which grade do you estimate receiving for this project? I expect/hope to recieve xxx	1.86***	0.73	$p < 0.01$	OC
15	If all grades of your class are added together, what do you think the average grade will be?	2.55***	0.55	$p < 0.01$	OC
16	Compared with other projects that you have completed during your education, how successful will you be with the current project?	3.51***	0.71	$p < 0.01$	OC
17	In comparison with you, how successful willl your class colleagues be on the current project?	2.97	0.68	$p = 0.59$	WC
13,17	Paired t-test question 13, 17, difference in mean = 0	0.00***		$p < 0.01$	OC

Note: OC: overconfident, WC: well-calibrated. The values in the column "Mean" refer to the mean of the answers given on a 5-part Likert-type scale. The theoretical median value is 3. Values are compared to the theoretical mean in the t-tests. The last line shows the results for a 2-sided t-test with mean=3 for the two questions 13 and 17.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Overconfidence in form of the better-than-average effect could be shown in 14 of the 17 analyzed topics in Table 6.3, of these, 13 are highly significant (p-Value < 0.01), and one is significant (p-Value < 0.1). Subjects are only well-calibrated in three of the analyzed topics, two of these items concern general education and the third item is a control item regarding expectations of the class mean.

6.3.2 Students Questionnaire 2

Reliability analysis Questionnaire 2 yields a Cronbach‘s $\alpha = 0.54$, which does not indicate particularly good internal reliability, but several of the subjects did not complete all items on the questionnaire, which degrades the quality of the returned questionnaires. Nevertheless, we are mainly interested in the comparison of the relevant values from the second questionnaire with the values from questionnaire 1, taken 2 months earlier. In order to make a longitudinal comparison, we remove the class data from questionnaire 1 whose participants are missing from questionnaire 2. This reduces the sample size in the amended questionnaire 1 to 90 final subjects and makes the samples comparable.

The second questionnaire was evaluated in comparison with the items from the reduced first questionnaire. In particular, the following items were contrasted:

Table 6.4: Longitudinal Questionnaire Comparison

Question	Qu. No.	Mean 1	Qu. No.	Mean 2	t-Value	p-Value
Estimated project time: Actual	1/18	80.10	2/16.2	41.52	2.43^{**}	0.017
Re-state time estimation	2/16.1	37.16	2/16.2	41.52	-1.79	0.079
Was project more/less intensive	2/8	3.38	2/8	3.00	3.35^{***}	0.001
Optimism/pessimism	1/4	3.87	2/4	3.95	-0.70	0.482
Grade expectation : Actual	1/14	1.89	2/10.2	1.90	0.34	0.731
Re-state grade expectation	2/10.1	1.89	2/10.2	1.84	0.53	0.598
Estim. techn. difficulty: Actual	1/19	2.589	2/13	3.20	-5.28^{***}	< 0.001
Confidence in techn. ability	1/20	3.72	2/14	3.83	-0.97	0.332

Note: The columns Qu. No. show the questionnaire numbers in the format questionnaire 1 or 2 / item number.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Professional Engineers

Without the open questions, Cronbach‘s $\alpha = 0.70$, which is a acceptable value for the internal reliability of the questionnaire items. The following results were obtained from t-tests of the items:

Table 6.5: Questionnaire Results for the Professional Engineers

Question No.	Mean	SD	t-Value	p-Value
Q4 Are you optimistic?	4.27	0.65	6.04***	< 0.001
Q6 How successful are you as an Engineer?	3.73	0.65	4.63***	< 0.001
Q7 Are you more or less diligent than your colleagues when planning and executing a project?	3.82	0.87	2.63**	0.022
Q8.1. Were large projects complex?	4.18	0.4	11.08***	< 0.001
Q8.2. Were medium-sized projects complex?	3.91	0.54	5.50***	< 0.001
Q17 How much do you learn from a typical project in planning?	3.82	1.25	2.13*	0.054
Q18 How much do you learn from a typical project in technical expertise?	4.45	0.69	7.98***	< 0.001
Q19 In comparison to your colleagues, do you learn more or less from new projects?	3.55	0.52	3.21***	0.008
Q20.4. Is your level of general education better or worse than your colleagues?	3.55	0.69	3.32***	0.006
Q23 How well do your clients rate your projects?	4.27	0.47	5.20***	< 0.001
Q24 Do you receive better or worse client feedback than your colleagues?	3.45	0.69	2.13*	0.054
Q25 How much does risk play a role when planning projects?	3.55	0.69	3.32***	0.006
Q27 Do you plan contingencies for risk?	3.45	1.21	1.86*	0.088

Note: The Table shows each question number in the leftmost column, the wording of each question in shown in the Appendix. The values in the column Mean and Std. Dev. show the arithmetic mean and standard deviation of the answers given to each question, and the p-Value shows a t-test significance of the difference between the actual mean and the theoretical mean of 3. t-tests were only performed for those questions that have a theoretical mean of 3.
*, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

In Table 6.5 only those questions are shown where significance in the answers is achieved. Optimism and self-professed success rate very highly. Overconfidence seems to be on a much lower level for the engineering professionals than for the students: the questions testing for overconfidence where comparisons with colleagues are asked for (Q7, Q19, Q20.4, and Q24) are all significant but not highly significant. Complexity of projects is obviously of major concern, this is coupled with significantly high results for risk awareness (Q25 and Q27). Learning from new projects also seems to feature highly for professional engineers. Summing up, there seems to be slight indications of overconfidence, but also risk awareness and the need for continual education can be shown from the answers.

6.4 Discussion

Overall, in the comparison of longitudinal data over a span of 8 weeks from questionnaire 1 to questionnaire 2, overconfidence and miscalibration was to be observed. Subjects were particularly susceptible to the better-than-average effect, a form of overconfidence, and were predictably overconfident when assessing their own abilities and time/effort necessary for the completion of an engineering project. Surprisingly, estimates of the grades were accurate, but this might be due to grading requirements or practices having been known in advance.

One of the main purposes of the study was to perform a longitudinal test for overconfidence by testing expectations before each subject was due to begin an individual engineering project, and to contrast the expectations with a) the actually achieved results (states by the subjects after the fact) and b) to re-test whether subjects accurately remembered the previously stated expectations or whether actually achieved results changed recollections of the previous estimates from 8 weeks before.

One questionnaire item asked subjects to estimate how much time they would be investing in the upcoming engineering project. The mean of the estimated time for the project before inception was 80.10 hours. On the second questionnaire, subjects were asked to state the number of hours actually invested, the mean was 41.52 hours. Obviously, a distinct indication of miscalibration is observable: The initially expected number of hours to be spent on project work is significantly (at $p < 5\%$) different from the number of hours actually invested, $t(106.14) = 2.43$, p-value = 0.01. The necessary number of hours to be spent on the project was obviously overestimated by 100% of the real value.

We re-tested this item in the second questionnaire and asked the subjects to restate what their estimation had been two months earlier. Interestingly, a distinct anchoring effect is observable: The first questionnaire had asked subjects to state their expectations of the number of hours that would be necessary for each subject to complete his or her individual engineering project work.

When asked to re-state the expectation of the number of hours necessary for project work that had been answered two months earlier (the mean was originally 80.10 hours), the mean of the new value was 37.16, and the mean of the answer to how many hours were actually spent on project work was 41.52. These two values are not significantly different from each other at either at either a 1% or 5% level, $t(50) = -1.79$, p-value =

0.08[23].

But the real indication of miscalibration and overconfidence in subjects' own estimations is the answer to the following item. The question was asked "was the effort required for the project more or less than what you had predicted when asked 2 months ago"? Subjects answered that the effort required was higher than what they had estimated 2 months ago, which was in total opposition to the answers actually provided, $t(59) = 3.35$, p-value = 0.001. So not only was the re-estimation biased by the anchoring effect, but additionally the subjects totally mis-remembered the estimations that they had originally provided and thought that the project had required more hours than they had forecast, when actually the opposite was the case. This shows distinct evidence of overconfidence in its form as miscalibration.

Optimism remained high and did not significantly change, $t(129.4) = -0.70$, p-value = 0.48. This was potentially due to the fact that the grading of the project was as favorable as had been expected by the subjects. The forecast accuracy was surprisingly accurate, even though the suspicion lends itself that there is a distinct anchoring effect in play. The originally forecast (mean) grade for the project was 1.89. The subjects had been asked in questionnaire 2 to re-state their original expectation. The (mean) value of 1.90 is almost exactly, $t(49) = 0.53$, p-value = 0.59, what had originally been forecast - the subjects were not prone to miscalibration in this case. However, as the stated achieved (mean) grade was 1.84, which is not significantly different from the originally forecast value, $t(89.44) = 0.34$, p-value = 0.73, the re-stated estimation could again have been anchored by the achieved value. Nevertheless, the forecast for the grade was surprisingly accurate. This could be for (at least) two reasons: 1) Either the students knew very well what to expect in regard to the particular teacher's grading or 2) forecasting for an important grade which was dependent on a calculable individual effort was feasible.

Despite the over-estimation of the necessary effort for the project, subjects under-estimated the degree of technical difficulty. The mean estimate of technical difficulty was 2.59, and the stated degree of difficulty after completion was 3.20, which is significantly higher, $t(125.72) = -5.28$, p-value < 0.001, and shows a high degree of over-confidence in the estimate of difficulty before beginning the project. Additionally, the subjects had been asked to provide an estimate of how confident they were that they had the necessary technical ability to complete the project. Interestingly, subjects significantly under-estimated their confidence of having the necessary technical ability for

[23]the null hypothesis is that the two values are equal, so a high p-value indicates that the null hypothesis of equality is not rejected.

the project, $t(123.74)$ = -0.97, p-value = 0.33. This would conform with the fact that good grades were obviously handed out for the project, but on the other hand objective technical difficulty had been under-estimated whereas confidence in own technical ability had also been under-estimated.

The professional engineers that completed their questionnaires had one common theme with the students: they scored similarly on the control questions regarding optimism and musicality[24]. The engineers (who at a mean of 50 years of age were a very experienced group) scored extremely high on the optimism scale, but according to their self-assessments, this optimism did not spill over into their self-estimates of risk-taking behavior, where they scored precisely the median.

Technical difficulties were assessed consistently; the complexity of projects increased according to the size of the projects. Add-ons to project time lines and costs were consistently assessed from experience, subjects stated that the risk premiums were adequate in almost 70% of cases, the average add-on was calculated at 8%, a value considered by all engineers to be slightly lower than that of peer group members (but not significantly so). Compared with the fact that most respondents stated that it could more often than not come to problems within projects, stated mis-calculations of projects cost an average of 19% in terms of time and 11% in terms of costs.

Obviously overconfidence in the validity of overrun buffers is more than eaten up by actually incurred overruns. The lesson to be learned here is that project risk management in terms of scope and cost needs to be accurately provided for.

Further evidence of overconfidence is visible in that subjects are fairly certain that they are more successful than their peer group, and that they consider themselves to be more diligent when planning and executing projects than other colleagues - a definite case of the better-than average effect.

It seems that the project- and risk management aspect is not held in such esteem as the technical aspects, even though the subjects are certified civil engineers and, as such, constant project managers. The polled engineers are of the opinion that they learn a little bit in terms of planning and costing with each new project, but a lot in technical development. At least they include learning effects from previous projects in

[24]The mean answers to the questionnaire item optimism were 3.95 and 4.27 for students and engineers respectively, both significantly scoring high on optimism. On the control qeustion regarding musicality, the mean answers were 2.83 and 2.82 for students and engineers, respectively. Both these values are significantly under the mean, indicating that both groups considered themselves to be un-musical

new ones, but it is not certain from the wording (and potential halo effect from the previous question) whether this might have meant to apply to technical development.

Clients are obviously very happy with results - most subjects record that clients are highly significantly happy (M=4.27) with the achieved results and the engineers state that each receives better feedback from clients than the peer group - a definite indication of the better-than-average effect.

6.5 Conclusion

Overconfidence can be detected everywhere - engineering students are prone to it in the same fashion as their elder colleagues with decades of experience. The engineering students tend to display age- and sex-related effects insofar that overconfidence in their own abilities tend to diminish over time. Students have improvement potential at both forecasting the time and effort for an engineering effort, as well as at remembering what they originally forecast - the combined behavioral biases of anchoring and better-than-average feelings can be significantly detected. Engineering students do not differ from other branches, as outlined in literature (see e.g. Lichtenstein et al., 1982), but the trend does seem to continue into professional life.

Experienced engineers seem to be adept at planning and executing their jobs and have reasonably well-calibrated opinions of themselves in many aspects versus their peer groups. However, somewhat surprisingly, the project risk management tasks seem not to have the same weight as the technical aspects. Considering that project risk is an inherent factor in all engineering tasks, increased attention to scope, time and cost risk management aspects are topics that could be addressed more in the engineering future. In order to improve results concerning risk management potential amongst engineers, however, a survey on a greater sample than the current one should be performed in future research.

6.6 Appendix

6.6.1 Fragebogen Schüler 1

Information

Bitte füllen Sie diesen Fragebogen in einem Durchgang aus. Ihre Antworten werden streng vertraulich behandelt und werden ausschließlich in anonymer Form für wissenschaftliche Zwecke im Rahmen einer Dissertation verwendet.

Bei Mehrfachantworten kreuzen Sie bitte immer nur ein Feld an, bei offenen Fragen bitte die Antwort (meist eine Zahl) in das dafür vorgesehene Kästchen schreiben. Zum Beispiel:

Ich bin

18	Jahre alt

oder: Ich bin

X	o
Weiblich	Männlich

Wir danken Ihnen für Ihre Mitarbeit an dieser Umfrage.

Allgemeiner Teil

Die Fragen in diesem Teil beziehen sich auf allgemeine Fähigkeiten oder Merkmale die prinzipiell auf Sie zutreffen könnten.

1) In welcher Klasse sind Sie? Ich bin in der Klasse

2) Wie alt sind Sie? Ich bin

	Jahre alt

3) Geben Sie Ihr Geschlecht an. Ich bin

o	o
Weiblich	Männlich

4) Würden Sie sich generell als eher optimistischen oder eher pessimistischen Menschen einschätzen? Ich bin meist

o	o	o	o	o
Sehr pessimistisch	Relativ pessimistisch	Weder noch	Relativ optimistisch	Sehr optimistisch

5) Erzielen Sie normalerweise bessere oder schlechtere Noten als Ihre KlassenkollegInnen? Meine Noten sind

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

6) Sind Sie der Meinung, daß Sie im Vergleich zu Ihren KlassenkollegInnen schulisch bessere oder schlechtere Leistungen erbringen? (egal ob sie damit bessere oder schlechtere Noten erzielen) Ich bin

o	o	o	o	o
Viel weniger erfolgreich	Etwas weniger erfolgreich	Gleich erfolgreich	Etwas erfolgreicher	Viel erfolgreicher

7) Sind Sie mehr oder weniger zufrieden mit Ihrem Gesamterfolg in der Schule? Ich bin

o	o	o	o	o
Überhaupt nicht zufrieden	Nicht besonders zufrieden	Egal	Recht zufrieden	Sehr zufrieden

8) Sind Sie im Vergleich zu Ihren KlassenkollegInnen mehr oder weniger sorgsam bei der Erstellung von Projektarbeiten? Ich bin

o	o	o	o	o
Viel weniger sorgsam	Etwas weniger sorgsam	Gleich sorgsam	Etwas sorgsamer	Viel sorgsamer

9) Sind Sie im Vergleich zu Ihren KlassenkollegInnen normalerweise schneller oder langsamer bei der Erstellung von Projektarbeiten? Ich bin

o	o	o	o	o
Viel langsamer	Etwas langsamer	Gleich schnell	Etwas schneller	Viel schneller

10) Glauben Sie, daß Sie zukünftig mehr oder weniger Gehalt bekommen werden als Ihre KlassenkollegInnen? Ich bekomme dann

o	o	o	o	o
Viel weniger	Etwas weniger	Gleich viel	Etwas mehr	Viel mehr

11) Sind Sie besser oder schlechter als Ihre KlassenkollegInnen in Bezug auf Ihre Fähigkeiten in den folgenden 5 Bereichen? Ich bin in Bezug auf mein/e:

11.1) Technischen Fähigkeiten (Rechnen, Analysieren, Zeichnen…)

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

11.2) Zeitmanagement

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

11.3) Kenntnis von Fremdsprachen

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

11.4) Allgemeinbildung

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

11.5) Musikalität

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

12) Ich erfülle im Vergleich zu meinen KlassenkollegInnen das Berufsbild eines idealtypischen Technikers

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

Spezieller Teil

Die Fragen in diesem Teil betreffen Ihre Fähigkeiten und Erwartungen bezüglich des kommenden Projekts.

13) Im Vergleich mit Ihren KlassenkollegInnen, wie gut werden Sie Ihr Projekt erledigen? Mein Projektergebnis wird

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

als das Projektergebnis der anderen sein.

14) Welche Note schätzen Sie, die Sie für dieses Projekt bekommen werden? Ich erwarte / hoffe, eine

o	o	o	o	o
5	4	3	2	1

bekommen.

15) Wenn man die Noten für dieses Projekt von allen SchülerInnen zusammenfasst, was könnte die Durchschnittsnote über alle gerechnet sein?

o	o	o	o	O
5	4	3	2	1

16) Verglichen mit anderen Projekten, die Sie persönlich im Laufe Ihrer Ausbildung gemacht haben, wie erfolgreich werden Sie im Vergleich dazu bei diesem konkreten Projekt sein? Ich werde

o	o	o	o	o
Viel weniger erfolgreich	Etwas weniger erfolgreich	Gleich erfolgreich	Etwas erfolgreicher	Viel erfolgreicher

sein.

17) Im Vergleich mit Ihnen, wie erfolgreich werden Ihre KollegInnen bei diesem Projekt sein? Sie werden

o	o	o	o	o
Viel weniger erfolgreich	Etwas weniger erfolgreich	Gleich erfolgreich	Etwas erfolgreicher	Viel erfolgreicher

sein.

18) Wieviel Gesamtaufwand werden Sie persönlich schätzungsweise in dieses Projekt in Stunden investieren? Ich habe vor, zirka

	Stunden

zu investieren.

19) Welchen technischen Schwierigkeitsgrad könnte dieses Projekt Ihrer Einschätzung nach aufweisen? Es könnte

o	o	o	o	o
Sehr schwierig	Relativ schwierig	Normal schwierig	Relativ leicht	Sehr leicht

werden.

20) Inwieweit haben Sie alle notwendigen technischen Fähigkeiten oder Fertigkeiten, um dieses Projekt bewältigen zu können? Ich habe

o	o	o	o	o
Keine	Wenige	Manche	Die meisten	Alle

dazu notwendigen technischen Fähigkeiten oder Fertigkeiten.

21) Wie oft haben Sie schon ähnliche Projekte selbst durchgeführt? Ich habe ähnliche Projekte schon

	Mal

gemacht.

22) Halten Sie das Projekt in der zur Verfügung stehenden Zeit für durchführbar? Ich habe für dieses Projekt

o	o	o	o	o
Viel zu wenig	Zu wenig	Ausreichend	Viel	Viel zu viel

Zeit zur Verfügung.

23) Wie wichtig ist dieses kommende Projekt für Sie? Das Projekt ist

o	o	o	o	o
Komplett unwichtig	Relativ unwichtig	Durchschnittlich wichtig	Relativ wichtig	Sehr wichtig

für mich.

6.6.2 Fragebogen Schüler 2

Information

Bitte füllen Sie diesen Fragebogen in einem Durchgang aus.

Ihre Antworten werden **streng vertraulich** behandelt und werden ausschließlich in anonymer Form für wissenschaftliche Zwecke im Rahmen einer Dissertation verwendet.

Bei **Mehrfachantworten** kreuzen Sie bitte immer nur **ein Feld** an.

Bei **offenen Fragen** bitte die Antwort (meist eine Zahl) in das dafür vorgesehene **Kästchen** schreiben.

Zum Beispiel:

Ich bin

18	Jahre alt

oder:

Ich bin

X	o
Weiblich	Männlich

Wir danken Ihnen für Ihre Mitarbeit an dieser Umfrage!

Allgemeiner Teil

Die Fragen in diesem Teil beziehen sich auf allgemeine Fähigkeiten oder Merkmale die prinzipiell auf Sie zutreffen könnten.

1) In welcher Klasse sind Sie? Ich bin in der Klasse

2) Wie alt sind Sie?

	Jahre

3) Geben Sie Ihr Geschlecht an.

o	o
Weiblich	Männlich

4) Würden Sie sich generell als eher optimistischen oder eher pessimistischen Menschen einschätzen? Ich bin meist

o	o	o	o	o
Sehr pessimistisch	Relativ pessimistisch	Weder noch	Relativ optimistisch	Sehr optimistisch

5) Erzielen Sie normalerweise bessere oder schlechtere Noten als Ihre KlassenkollegInnen? Meine Noten sind

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

6) Sind Sie im Vergleich zu Ihren KlassenkollegInnen mehr oder weniger sorgsam bei der Erstellung von Projektarbeiten? Ich bin

o	o	o	o	o
Viel weniger sorgsam	Etwas weniger sorgsam	Gleich sorgsam	Etwas sorgsamer	Viel sorgsamer

7) Sind Sie im Vergleich zu Ihren KlassenkollegInnen normalerweise schneller oder langsamer bei der Erstellung von Projektarbeiten? Ich bin

o	o	o	o	o
Viel langsamer	Etwas langsamer	Gleich schnell	Etwas schneller	Viel schneller

Spezieller Teil

Die Fragen in diesem Teil betreffen Ihre Fähigkeiten und Erwartungen bezüglich des aktuell abgeschlossenen Projektes.

8) Der tatsächliche Zeitaufwand für dieses Projekt war

o	o	o	o	o
Viel niedriger	Etwas niedriger	Gleich hoch	Etwas höher	Viel höher

als der vermutete/geschätzte Zeitaufwand vor Projektstart.

9) Im Vergleich mit Ihren KlassenkollegInnen, wie gut haben Sie Ihr Projekt erledigt? Mein Projektergebnis war

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

als das Projektergebnis der anderen.

10.1.) Welche Note schätzen Sie, die Sie für dieses Projekt bekommen werden? Ich erwarte / hoffe, eine

o	o	o	o	o
1	2	3	4	5

zu bekommen.

10.2.) Falls Sie bereits jetzt eine Note für dieses Projekt erhalten haben: Welche Note haben Sie erhalten?

o	o	o	o	o
1	2	3	4	5

11) Verglichen mit anderen Projekten, die Sie persönlich im Laufe Ihrer Ausbildung gemacht haben, wie erfolgreich waren Sie im Vergleich dazu bei diesem konkreten Projekt? Ich war bei diesem Projekt

o	o	o	o	o
Viel weniger erfolgreich	Etwas weniger erfolgreich	Gleich erfolgreich	Etwas erfolgreicher	Viel erfolgreicher

12) Im Vergleich mit Ihnen, wie erfolgreich waren Ihre KollegInnen bei diesem Projekt? Meine KollegInnen waren

o	o	o	o	o
Viel weniger erfolgreich	Etwas weniger erfolgreich	Gleich erfolgreich	Etwas erfolgreicher	Viel erfolgreicher

13) Wie war der tatsächliche technische Schwierigkeitsgrad bei diesem Projekt Ihrer Einschätzung nach im Vergleich zum vermuteten Schwierigkeitsgrad? Der technische Schwierigkeitsgrad war tatsächlich

o	o	o	o	o
Viel leichter	Leichter	Gleich hoch	Schwieriger	Viel schwieriger

als vermutet/geplant.

14) Inwieweit hatten Sie alle notwendigen technischen Fähigkeiten oder Fertigkeiten, um dieses Projekt bewältigen zu können? Ich hatte

o	o	o	o	o
Keine	Wenige	Manche	Die meisten	Alle

dazu notwendigen technischen Fähigkeiten oder Fertigkeiten.

15) Wie oft haben Sie schon ähnliche Projekte selbst durchgeführt?

o	o	o	o
Noch nie	1 – 3 Mal	4 – 10 Mal	Über 10 Mal

16.1.) Wieviel Gesamtaufwand (in Stunden) hatten Sie ursprünglich <u>geplant</u> in dieses Projekt zu investieren?

	Stunden

17.2.) Wieviel Gesamtaufwand (in Stunden) haben Sie <u>tatsächlich</u> in dieses Projekt investiert?

	Stunden

6.6.3 Fragebogen Techniker

Information

Bitte füllen Sie diesen Fragebogen in einem Durchgang aus. Ihre Antworten werden streng vertraulich behandelt und werden ausschließlich in anonymer Form für wissenschaftliche Zwecke im Rahmen einer Dissertation verwendet.

Bei Mehrfachantworten kreuzen Sie bitte immer nur ein Feld an, bei offenen Fragen bitte die Antwort (meist eine Zahl) in das dafür vorgesehene Kästchen schreiben. Zum Beispiel:

Ich bin

45	Jahre alt

oder: Ich bin

X	o
Weiblich	Männlich

Wir danken Ihnen für Ihre Mitarbeit an dieser Umfrage.

Allgemeiner Teil

Die Fragen in diesem Teil beziehen sich auf allgemeine Fähigkeiten oder Merkmale die prinzipiell auf Sie zutreffen könnten.

1) Wie alt sind Sie? Ich bin

	Jahre alt

2) Geben Sie bitte Ihr Geschlecht an. Ich bin

o	o
Weiblich	Männlich

3) Wieviele Jahre sind Sie schon berufstätig ? Ich habe

	Jahre Berufserfahrung

4) Würden Sie sich generell als eher optimistischen oder eher pessimistischen Menschen einschätzen? Ich bin meist

o	o	o	o	o
Sehr pessimistisch	Relativ pessimistisch	Weder noch	Relativ optimistisch	Sehr optimistisch

5) Würden Sie sich generell als eher risikofreudigen oder eher risikoaversen Menschen einschätzen? Ich bin meist

o	o	o	o	o
Gar nicht risikofreudig	Relativ wenig risikofreudig	Neutral	Relativ risikofreudig	Sehr risikofreudig

Spezieller Teil

Die Fragen in diesem Teil betreffen Ihre Fähigkeiten und Erwartungen in Ihrem Beruf als Techniker.

6) Wie erfolgreich sind Sie Ihrer Einschätzung nach in Ihrem Beruf als Techniker?

o	o	o	o	o
Nicht erfolgreich	Wenig erfolgreich	Neutral	Relativ erfolgreich	Sehr erfolgreich

7) Würden Sie sich im Vergleich zu Ihren KollegInnen bei der Planung und Durchführung von Projekten als mehr oder weniger sorgfältig einschätzen?

o	o	o	o	o
Viel weniger	Etwas weniger	Gleich viel	Etwas mehr	Viel mehr

2) Wieviele Projekte haben Sie schon durchgeführt ? Ich habe

	Große Projekte
	Mittlere Projekte
	Kleine Projekte

durchgeführt.

8) Waren diese Projekte vom technischen Schwierigkeitsgrad und der Komplexität her als mehr oder weniger schwierig/komplex einzustufen?

8.1) Die großen Projekte waren

o	o	o	o	o
Sehr einfach	Relativ einfach	Neutral	Relativ komplex	Sehr komplex

8.2) Die mittleren Projekte waren

o	o	o	o	o
Sehr einfach	Relativ einfach	Neutral	Relativ komplex	Sehr komplex

8.3) Die kleinen Projekte waren

o	o	o	o	o
Sehr einfach	Relativ einfach	Neutral	Relativ komplex	Sehr komplex

9) Führen Sie oft vom Aufgabengebiet und der technischen Komplexität ähnliche Projekte durch oder eher unterschiedliche?

o	o	o	o	o
Sehr unterschiedlich	Relativ unterschiedlich	Neutral	Relativ ähnlich	Sehr ähnlich

10) Wieviel Zeit (als Prozentsatz des Gesamtaufwands des Projekts) verbringen Sie mit der Planung der Aufwände, Zeiten und Kosten für das Projekt?

	% des Gesamtaufwands

11) Wenn Sie nachher das Projekt finalisiert haben und rückblicken: um wieviel Prozent haben Sie sich dabei im Schnitt beim Zeitaufwand verschätzt, den Sie für das Projekt benötigen werden?

	% des Zeitaufwands

12) Um wieviel Prozent verschätzten Sie sich dabei im Schnitt bei den voranschlagten Kosten?

	% der Kosten

13) Im Vergleich zu Ihren KollegInnen, verschätzen Sie sich mehr oder weniger als sie bei den voranschlagten Kosten?

o	o	o	o	o
Viel weniger	Etwas weniger	Gleich viel	Etwas mehr	Viel mehr

14) Haben Sie sich schon mal bei Projekten so grob verschätzt, daß das Projekt in der geplanten Form undurchführbar war und Sie mit der Planung neu beginnen mußten?

o	o	o	o	o
Nie	Selten	Manchmal	Oft	Sehr oft

15) Haben schon mal bei Ihren Projekten unentdeckte Fehlplanungen oder –konstruktionen später zu größeren Schäden geführt?

o	o	o	o	o
Nie	Selten	Manchmal	Oft	Sehr oft

16) Ihrer Einschätzung nach, bei wieviel Prozent der Projekte Ihrer KollegInnen haben unentdeckte Fehlplanungen oder –konstruktionen später zu größeren Schäden geführt?

	% der Projekte

17) Wieviel lernen Sie aus einem typischen Projekt **bei der Planung der Zeit, Aufwände und Kosten** an Erfahrungwerten dazu?

o	o	o	o	o
Nichts	Relativ wenig	Neutral	Etwas	Viel

18) Wieviel lernen Sie aus einem typischen Projekt **aus technischer Sicht** an Erfahrungswerten dazu?

o	o	o	o	o
Nichts	Relativ wenig	Neutral	Etwas	Viel

19) Im Vergleich zu Ihren KollegInnen, inkludieren Sie Lerneffekte aus vergangenen Projekten mehr oder weniger in Ihre neuen Projekte?

o	o	o	o	o
Viel weniger	Etwas weniger	Gleich viel	Etwas mehr	Viel mehr

20) Sind Sie besser oder schlechter als Ihre KollegInnen in Bezug auf Ihre Fähigkeiten in den folgenden 5 Bereichen? Ich bin in Bezug auf mein/e:

20.1) Technischen Fähigkeiten (Rechnen, Analysieren, Zeichnen…)

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

20.2) Zeitmanagement

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

20.3) Kenntnis von Fremdsprachen

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

20.4) Allgemeinbildung

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

20.5) Musikalität

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

21) Ich erfülle im Vergleich zu meinen KollegInnen das Berufsbild eines idealtypischen Technikers

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich gut	Etwas besser	Viel besser

22) Ist Ihr Jahreseinkommen Ihrer Einschätzung nach im Vergleich zu Ihren KollegInnen eher höher oder niedriger?

o	o	o	o	o
Viel niedriger	Etwas niedriger	Gleich	Etwas höher	Viel höher

23) Wie gut bewerten Ihre Kunden im Schnitt Ihre Projektarbeit?

o	o	o	o	o
Sehr schlecht	Schlecht	Neutral	Gut	Sehr gut

24) Bekommen Sie Ihrer Einschätzung nach im Vergleich zu Ihren KollegInnen besseres oder schlechteres Feedback von den Kunden?

o	o	o	o	o
Viel schlechter	Etwas schlechter	Gleich	Etwas besser	Viel besser

25) Inwieweit spielt das Risiko, daß es im Laufe eines Ihrer Projekte zu Problemen kommen kann bei Ihrer Projektplanung eine Rolle?

o	o	o	o	o
Gar nicht	Selten	Neutral	Manchmal	Oft

26) Betreiben Sie bei Ihrer Projektplanung Vorsorge zu Risikomanagementzwecken in technischer Hinsicht (zB Überdimensionierung, Materialpuffer oder dergleichen)?

o	o	o	o	o
Gar nicht	Selten	Neutral	Manchmal	Oft

27) Betreiben Sie bei Ihrer Projektplanung Vorsorge zu Risikomanagementzwecken in planerischer Hinsicht (zB Zeitpuffer, Aufwandspuffer oder dergleichen)?

o	o	o	o	o
Gar nicht	Selten	Neutral	Manchmal	Oft

28) Wieviel Prozent Ihrer Kostenschätzung für ein Projekt ist im Schnitt als Risikoaufschlag oder Risikovorsorge vorgesehen?

	% der Kostenschätzung

29) Wie berechnen Sie diese Risikoaufschläge oder Risikovorsorgen?

o	o	o	o	o
Intuitiv	Aus Erfahrungswerten	Laut externen Vorgaben	Jedesmal neu quantifiziert	Anders

30) In wieviel Prozent Ihrer Projekte reicht dieser Risikoaufschlag für die Abdeckung potentiell auftretender Umstände aus?

	% der Projekte

31) Ihrer Einschätzung nach berechnen Sie beim Risikoaufschlag / Risikovorsorge mehr oder weniger als Ihre KollegInnen?

o	o	o	o	o
Viel weniger	Etwas weniger	Gleich viel	Etwas mehr	Viel mehr

Chapter 7

Integrating Behavioral Aspects into Risk Management

Even though the risk management recommendations in this chapter deal with behavioral biases in finance, the principles are valid for all forms of risk or quality management or control. In chapter 6 we provided the results of tests for overconfidence outside the financial world. These results show that recognition and management of behavioral biases is not restricted to finance, rather tests for overconfidence in engineering students and professionals show that behavioral biases are just as prevalent in a technical field as in finance.

As mentioned in chapter 1, the field of finance has started to incorporate findings from psychology dealing with cognitive biases, mainly since Simon (1955) and Tversky and Kahneman (1974), resulting in a new field called behavioral finance (see Mullainathan and Thaler, 2000). Behavioral finance has come a long way since the seminal articles quoted above, but has not yet found its way into risk management within academia. This chapter uses the findings from the previous chapters, puts them in a risk management context and provides recommendations of how to implement the findings into risk management practice.

7.1 Risk Management Objectives

Supervisory authorities have spent quite some time evaluating and devising new governance regulations (also due to the failures and excesses in recent times outlined in chapter 1). Risk management principles must be set and supervised by the board of an institution (see e.g. Hofmann, 2008). Risk management processes can be quite complex, but at the strategic level are simple - the choices for risk management are (see Crouhy et al., 2000):

1. Avoid risk by choosing not to undertake some activities.
2. Transfer risk to third parties through insurance, hedging, or outsourcing.
3. Mitigate risk through preventive and detective control measures.
4. Accept risk, recognizing that undertaking certain risky activities generate shareholder value.

To fulfill its risk governance responsibilities, the board must ensure that the institution has put an effective risk management program in place that is consistent with its choices of fundamental strategy and risk appetite. This includes implementing procedures for all types of risk, such as business risk, market risk, credit risk, liquidity risk, operational risk, legal risk, etc. and making sure that all appropriate policies, methodologies and infrastructure are set in place. These would include but not be limited to:

- Risk decision committees
- Limits in different risk categories
- Internal audit functionalities
- Documented processes, handbooks and policies

The recommendations outlined in this chapter deal mainly with the risk category of market risk (the risk of adverse movements in interest rates, foreign exchange rates, stock prices and commodity prices). This is due to the fact that the research in the previous chapters was focused on the detection of the effects of behavioral biases within a market risk context. However, the findings are just as applicable to other risk categories, in particular also the detection of behavioral biases could be assigned to the purview of an operational risk department.

7.2 The Risk Management Process

There exist several versions of a risk management process cycle, but most are similar and agree that risk management can only be displayed as a feedback cycle. The Project Management Institute defines the risk management process with the following steps: 1) risk management planning, 2) risk identification, 3) qualitative risk analysis, 4) quantitative risk analysis, 5) risk response planning, and 6) risk monitoring and control (Project Management Institute, 2004). These steps take place in a recurring feedback cycle, see Figure 7.1. This particular cycle was obviously developed for project management, but the fundamental steps in the cycle (risk identification, risk analysis/quantification, risk response planning and risk monitoring are those steps that accurately describe the usual risk management process (with the potential addition of risk reporting to the outlined steps).

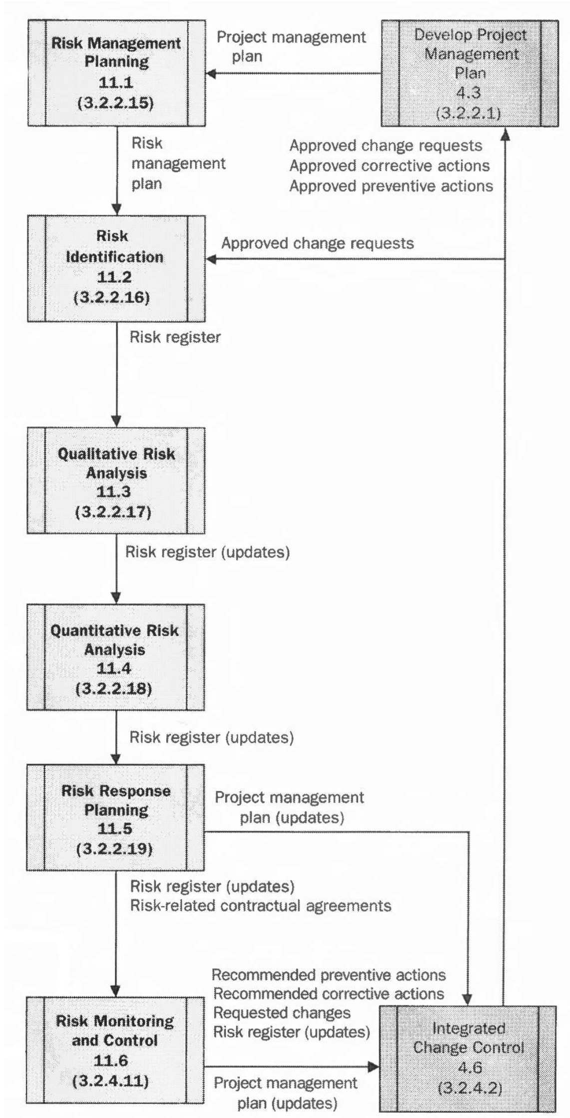
In 2004, the Committee of Sponsoring Organizations of the Treadway Commission (COSO) published a framework that had been previously developed to effectively identify, assess, and manage risk. Its purpose was to outline "*[f]our components [that] relate to the design and operation of the system of internal control: control environment, risk assessment, control activities, and information and communication. The fifth component — monitoring — is designed to ensure that internal control continues to operate effectively*" (of Sponsoring Organizations of the Treadway Commission, 2004).

Since then it has been revisited and enhanced, and re-issued as the COSO II framework, "*[...] which further developed the understanding of how all five internal control components work cohesively to form an effective internal control system*" in 20 principles (of Sponsoring Organizations of the Treadway Commission, 2008). The COSO II Integrated Framework for risk management is shown in Figure 7.2.

The inclusion of new risk management procedures has to take place within the context of existing procedures, as expanding risk management to include behavioral aspects is only an extension of existing frameworks. Current frameworks, e.g. in a market risk context, identify the different financial exposures to market risk factors, structure them into risk categories, such as interest rates, foreign exchange exposures etc., measure them with models such as Value at Risk, limit the traders' potential exposures via the models, report the figures, and integrate the new data from the present day into the next day's process in the form of a feedback cylce.

This procedure must necessarily be followed when including behavioral aspects in order to adhere to existing structures and processes that function within an institution.

Figure 7.1: The PMBOK Risk Management Cycle



Source: The Project Management Institute, 2004, p. 241

The behavioral aspects suggested in this dissertation do not constitute new process frameworks, but provide new aspects within existing frameworks. What is new, is that the behavioral aspects need to be added to and integrated into the same procedure, i.e. exposures to biases need to be identified, assigned to a category (structured), measured, limited, reported and included into the feedback cycle. These steps can be part of the COSO structure consisting of the Control Environment, Risk Assessment, Control Activities, Information & Communication, and Monitoring (see Figure 7.2). Where the behavioral aspects outlined in the previous chapters of this dissertation should be integrated into existing frameworks is outlined in the following sections.

Figure 7.2: The COSO II Integrated Framework

Source: The Committee of Sponsoring Organizations of the Treadway Commission, 2008, p. 1

7.3 Early Warning of Biases

7.3.1 Management of Overconfidence

In chapter 4 (Forecasting Overconfidence), we identified that recognizing those traders who would later prove to be overconfident (and thus in an institutional trading setting, cause losses for their employer) is possible via the methods outlined in that chapter. Overconfidence can only be identified by collecting information about people's inclinations ex-ante and comparing these answers with the then realized results ex-post. This was performed in the conducted experiment via psychological questionnaires containing carefully selected items.

In order to recognize the overconfidence of traders, this information has to be collected by the employer (the institition) on a regular basis, presumably via the human resource (HR) departement by immediate superiors in the form of intra-yearly performance reviews. Crafting such questionnaires is an art unto itself. In chapter 3 we found that the most important items related to overconfidence were subjects' self-professed prediction abilities and the importance of the subject matter to the person. This information must be collected via carefully crafted psychological questionnaires, and changed regularly so that employees are not able to "arbitrage" the questions or fake the results.

There is some existing research in the field of psychology on the topic of sensation seeking via questionnaires which is relevant to this topic (see Hoyle et al., 2002), but these questionnaires were developed with risk-seekers outside of finance in mind. Questionnaires for trading or investment purposes would have to be designed explicitly for the purpose, including feedback from HR departments and potentially from unions, would have to comply with legal requirements and would have to be carefully tested in pilot environments within a relevant institution. This is a topic for further research.

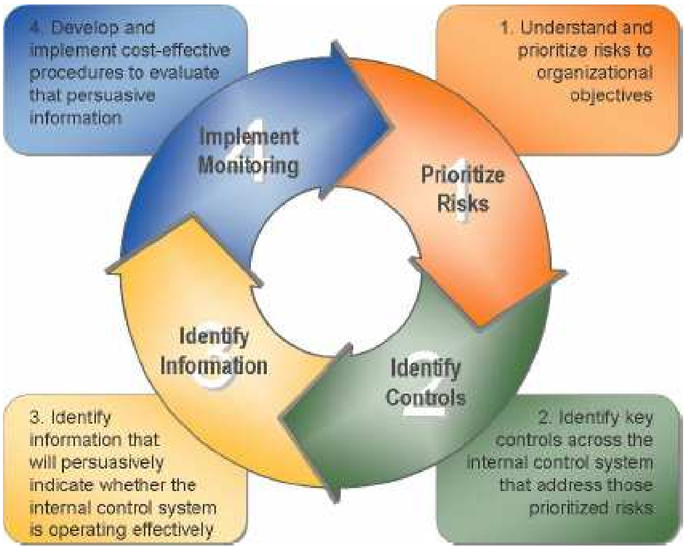
The risk management process for early recognition of overconfidence in traders could consist of several distinct steps contained within the blocks Control Environment, Risk Assessment, and Control Activities in the COSO framework (see Figure 7.2):

1. **Step:** Collection of the necessary information via questionnaires (risk identification / control environment).
2. **Step:** Analytical identification of potentially overconfident traders (risk quantification / risk assessment).
3. **Step:** Selection of those traders that potentially pose a problem (risk response planning / control. activities). As, potentially, many traders could prove to be overconfident at any given point in time, care should be taken to only select the most overconfident of these, the objective should be to catch the top most overconfident percentile (whichever is to be chosen) who would generate substantial losses.
4. **Step:** Devising an individualized limit concept for applying limits to those traders who are within the most overconfident percentile in a given period (risk monitoring and control / control environment).
5. **Step:** Constant monitoring of the behavior of traders with the implemented models (risk monitoring and control / monitoring).
6. **Step:** Integrating the lessons learned from one period into the next (control activities).
7. **Step:** Reporting the limits, losses, and successes from the individualized limit and measurement concept (information and communication).

According to the COSO II framework (of Sponsoring Organizations of the Treadway Commission, 2008), the monitoring design of the Framework includes the following formalized steps that incorporate the activities outlined above:

- Prioritizing the risks,
- Identifying the correct controls,
- Identifying the necessary information, and
- Implementing the monitoring framework.

Figure 7.3: The COSO II Monitoring Process



Source: The Committee of Sponsoring Organizations of the Treadway Commission, 2008, p. 18

The COSO II process, shown in Figure 7.3, is the guideline for the recommendations outlined above for measuring, limiting and monitoring behavioral biases. The outlined steps can all be set within the monitoring process.

7.3.2 Management of the Disposition Effect

In chapter 5 we tested for evidence of the disposition effect on an experimental market and a real market. It was shown that even though professional traders and investors are less prone to the disposition effect than retail investors or amateurs, the effect is still measurable, and that it occurs in differentiated form.

The disposition effect needs to be managed differently from the overconfidence bias, as both the necessary data and the monitoring process are different. The disposition effect does not require gathering ex-ante data. Rather, it is an ongoing process that can be measured and monitored frequently (daily, if necessary). Steps in the risk management process for the disposition effect could include:

1. **Step:** Definition of the portfolio and trader data (control environment).

2. **Step:** Definition of the analytics for calculating the disposition effect (control environment).

3. **Step:** Performing of daily disposition effect calculations (risk assessment).

4. **Step:** Evaluation of the top percentile of the trades who are prone to the dipsosition effect (risk assessment).

5. **Step:** Creation of individual stop-loss limits for these traders and regular adaptation of these limits (control environment).

6. **Step:** Regular monitoring of the individual stop-loss limits (monitoring).

7. **Step:** Regular reporting of the individualized limits, the limit excesses and limit utilization (information and communication).

8. **Step:** Inclusion of the data from previous periods into the next period as a feedback cycle (control activities).

7.4 Conclusion

The outlined methods that could be implemented within a risk management context are only the beginning of a new form of individualized risk management based on behavioral aspects. As outlined in the introduction to this chapter, risk management and cognitive biases such as overconfidence are now household words in news articles. The usage of techniques to recognize, monitor and control should provide an additional tool in a risk manager's toolbox. In this dissertation we have analyzed only two of the more than 90 known cognitive biases that are currently cited by the online platform Wikipedia (2011). Further research should branch out and include risk management techniques and methodologies for other biases, as well as implementing the risk management basics for behavioral risk management in practice. The risk management practices outlined here can be transferred to other, non-finance areas of risk or quality control. As behavioral finance has become a mainstream branch of finance, it is to be hoped that behavioral risk management could be as successful.

References

Allport, G. W. (1937). *Personality: A psychological interpretation*. Holt, Rhinehart & Winston, New York.

Allport, G. W. (1966). Traits revisited. *American Psychologist*, 21:1–10.

Allport, G. W. and Odbert, H. S. (1936). Trait names, a psycholexical study. *Psychological Monographs, 47*, 1:Whole 211.

Altman, M. (2010). Prospect theory and behavioral finance. In Baker, H. K. and Nofsinger, J. R., editors, *Behavioral Finance: Investors, Corporations, and Markets*. John Wiley & Sons, Inc.

Andersson, P., Edman, J., and Ekman, M. (2005). Predicting the world cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21:565–576.

Ariely, D. (2008). *Predictably Irrational: the hidden forces that shape our decisions*. HarperCollinsPublishers.

Armstrong, J. S. (2002). *Principles of Forecasting*. Kluwer Academic Publishers.

Aronson, E., Wilson, T. D., and Akert, R. M. (2010). *Social Psychology*. Pearson Education, Inc., 7th edition.

Arrow, K. J. (1986). Rationality of self and others in an economic system. *Journal of Business*, 59:385–399.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193(5):31–35.

Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'É.N.S.*, 17:21–86.

Ball, R. and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2):159–178.

Bao, J., Pan, J., and Wang, J. (2011). The illiquidity of corporate bonds. *The Journal of Finance*, LXVI(3):911–946.

Barber, B. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, pages 261–292.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., and Odean, T. (2007). Is the aggregate investor reluctant to realise losses? evidence from taiwan. *European Financial Management*, 13(3):423–447.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, LV(2).

Barberis, N. and Huang, M. (2001). Mental accounting, loss aversion, and individual stock returns. *The Journal of Finance*, LVI(4):1247–1292.

Barberis, N. and Xiong, W. (2009). What drives the disposition effect? ana analysis of a long-standing preference-based explanation. *The Journal of Finance*, LXIV(2):751–784.

Bardsley, N., Cubett, R., Loomes, G., Moffatt, P., Starmer, C., and Sugden, R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

BBC (2011a). Japan earthquake: Explosion at fukushima nuclear plant. online.

BBC (2011b). Japan earthquake: Tsunami hits north-east. online.

Belsky, G. and Gilovich, T. (1999). *Why Smart People Make Big Money Mistakes*. Simon & Schuster, New York.

Ben-David, I., Graham, J. R., and Harvey, C. R. (2007). Managerial overconfidence and corporate policies. *NBER Working Paper 13711*.

Benos, A. (1998). Aggressiveness and survival of overconfident traders. *Journal of Financial Markets*, 1(3-4):353–383.

Bentham, J. (1781). *An Introduction to the Principles of Morals and Legislation*, chapter I: Of the Principle of Utility. Oxford at the Clarendon Press.

Berg, J., Forsythe, R., Nelson, F., and Rietz, T. (2000). Results from a dozen years of election futures markets research. Technical report, College of Business Administration, University of Iowa.

Berg, J. and Rietz, T. (2003). Prediction markets as decision support systems. *Information Systems Frontiers*, 5(1):79–93.

Berg, J. E., Nelson, F. D., and Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, 24:285–300.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, V:175–192.

Bernstein, P. L. (2005). *Capital Ideas*. John Wiley & Sons, Inc.

Bhattacharya, S. and Haslett, J. (2007). Importance re-sampling mcmc for cross-validation in inverse problems. *Bayesian Analysis*, 2(2):385–408.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Black, F. (1986). Noise. *The Journal of Finance*, XLI(3):529–543.

Borge, D. (2001). *The Book of Risk*. John Wiley & Sons, Inc.

Borooah, V. K. (2002). *Logit and Probit: ordered and multinomial models*. Sage University Paper series on Quantitative Applications in the Social Sciences 138.

Brooks, C. (2008). *Introductory Economometrics for Finance*. Cambridge University Press.

Camerer, C. and Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874.

Campbell, J. and Mankiw, N. G. (1987). Are output fluctuations transitory? *Quarterly Journal of Economics*, 102:857–880.

Carmona, R. A. (2004). *Statistical Analysis of Financial Data in S-Plus*. Springer.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1:245–276.

cbcnews (2011). Ubs case latest failure of risk management. online.

Chamberlin, E. H. (1948). An experimental imperfect market. *The Journal of Political Economy*, 56(2):95–108.

Chen, G., Kim, K. A., Nofsinger, J. R., and Rui, O. M. (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making*, 20:425–451.

Choe, H. and Eom, Y. (2009). The disposition effect and investment performance in the futures market. *The Journal of Futures Markets*, 29(6):496–522.

Chou, R. K. and Wang, Y.-Y. (2011). A test of the different implications of the overconfidence and disposition hypotheses. *Journal of Banking & Finance*, 35:2037–2046.

Clausen, S.-E. (1998). *Applied Correspondence Analysis: An Introduction*. 07-121. Sage University Papers Series on Quantitative Applications in the Social Sciences.

Cochrane, J. H. (1988). How big is the random walk in gnp? *Journal of Political Economy*, 96:893–920.

Cochrane, J. H. (1991). A critique of the application of unit root tests. *Journal of Economic Dynamics*, 15:275–284.

Cohen, Y. and Cohen, J. (2008). *Statistics and Data with R*. John Wiley & Sons, Inc.

Commission, S. F. B. (2008). Subprime crisis: Sfbc investigation into the causes of the write-downs of ubs ag. *Swiss Federal Banking Commission*.

Constantinides, G. M. (1983). Capital market equilibrium with personal tax. *Econometrica*, 51:611–636.

Cournot, A. A. (1838). *Researches on the Mathematical Principles of the Theory of Wealth*.

Courtault, J.-M. (2000). Louis bachelier: On the centenary of theorie de la speculation. *Mathematical Finance*, 10(3):341–353.

Coval, J. D. and Shumway, T. (2005). Do behavioral biases affect prices? *The Journal of Finance*, IX(1):1–34.

Cowgill, B., Wolfers, J., and Zitzewitz, E. (2009). Using prediction markets to track information flows: Evidence from google. In Das, S., Ostrovsky, M., Pennock, D., and Szymanski, B. K., editors, *Auctions, Market Mechanisms and Their Applications*, volume 14. Springer, Berlin.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, 1:309–324.

Cowles, A. (1944). Stock market forecasting. *Econometrica*, 12:206.

Cowpertwait, P. S. P. and Metcalfe, A. V. (2009). *Introductory Time Series with R.* Springer.

Cromwell, J. B., c. Labys, W., and Terraza, M. (1994). *Univariate Tests for Time Series Models.* Sage University Paper series on Quantitative Applications in the Social Sciences.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Crouhy, M., Galai, D., and Mark, R. (2000). *Risk Management.* McGraw-Hill Professional.

Cryer, J. D. and Chan, K.-S. (2008). *Time Series Analysis with Applications in R.* Springer.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance*, LVI(3):921–965.

Davidson, R. and MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 49:781–793.

DeBondt, W. F. and Thaler, R. (1984). Does the stock market overreact? *The Journal of Finance*, 40(3):793–805.

DeFinetti, B. (1937). "la prévision: Ses lois logiques, ses sources subjectives", annales de l"institut henri poincaré, 7: 1–68; translated as "foresight. its logical laws, its subjective sources". In H. E. Kyburg, J. and Smokler, H. E., editors, *Studies in Subjective Probability.* Robert E. Krieger Publishing Company, 1980.

DeWeaver, M. A. and Shannon, R. (2010). Waning vigilance and the disposition effect: Evidence from thailand on individual investor decision making. *The Journal of Socio-Economics*, 39:18–23.

Dhar, R. and Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 52(5):726–740.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(365):427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057–1072.

Dictionaries, O. (2010). *Oxford Dictionaries.* Oxford University Press.

Doukas, J. and Petmezas, D. (2007). Acquisitions, overconfident managers and self-attribution bias. *European Financial Management*, 13(3):531–577.

Dowie, J. (1976). On the efficiency and equity of betting markets. *Economica*, 43:139–150.

Drucker, P. F. (1974). *Management.* Harper & Row.

Dumas, J. (1999). Usability testing methods: Subjective measures, part ii - measuring attitudes and opinions. *American Institutes for Research*.

Dunteman, G. H. (1989). *Principal Components Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. *Biometrika*, 38:159–171.

Eckholm, A. and Pasternack, D. (2007). Overconfidence and investor size. *European Financial Management*, 14(1):82–98.

Eysenck, H. J. (1990). Biological dimensions of personality. In Pervin, L. A., editor, *Handbook of personality theory and research*, pages 244–276. Guilford Press, New York.

Eysenck, M. (1998). *Psychology: An Integrated Approach*. Addison Wesley Longman.

Fabozzi, F. J. and Modigliani, F. (2003). *Capital Markets: Institutions and Instruments*. Prentice Hall, 3rd edition.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.

Fama, E. (1998). Efficiency survives the attack of the anomalies. *GSB Chicago*, pages 14–16.

Fama, E. F. and French, K. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96:246–273.

Festinger, L. A. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fischhoff, B. (1982). Debiasing. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Fisher, G. (2008). Euro 2008 in quotes. web.

Fiske, S. T. and Taylor, S. E. (2007). *Social Cognition: From Brains to Culture*. McGraw-Hill Professional.

Forsythe, R., Frank, M., Krishnamurthy, V., and Ross, T. (1995). Using market prices to predict election results: The 1993 ubc election stock market. *Canadian Journal of Economics*, 28:770–793.

Forsythe, R., Nelson, F., Neumann, G., and Wright, J. (1992). Anatomy of an experimental political stock market. *The American Economic Review*, 82(5):1142–1161.

Forsythe, R., Rietz, T. A., and Ross, T. W. (1999). Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior & Organization*, 39:83–110.

Friedman, D. and Rust, J. (1993). *The Double Auction Market: Institutions, Theories, and Evidence*. Perseus Books.

Frino, A., Johnstone, D., and Zheng, H. (2004). The propensity for local traders in futures markets to ride losses: Evidence of irrational or rational behavior? *Journal of Banking & Finance*, 28:353–372.

Galton, F. (1907). Vox populi. *Nature*, 75:450–451.

Garvey, R. and Murphy, A. (2004). Are professional traders too slow to realize their losses? *Financial Analysts Journal*, 60(4):35–43.

George, D. and Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* Boston: Allyn & Bacon, 11.0 update (4th ed.) edition.

Gerrig, R. J. and Zimbardo, P. G. (2010). *Psychology and Life*. Allyn & Bacon, 19th edition.

Gervais, S. and Odean, T. (2001). Learning to be overconfident. *The Review of Financial Studies*, 14(1):1–27.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1):20–29.

Gigerenzer, G. and Selten, R. (1999). *Bounded Rationality The Adaptive Toolbox*. Dahlem Workshop Reports.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(15):900–901.

Glaser, M., Langer, T., and Weber, M. (2007). On the trend recognition and forecasting ability of professional traders. *Decision Analysis*, 4(4):176–193.

Gorsuch, R. L. (1997). Explanatory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68:532–560.

Gort, C. (2009). Overconfidence and active management: An empirical study across swiss pension plans. *The Journal of Behavioral Finance*, 10:69–80.

Greenacre, M. (2007). *Correspondence Analysis in Practice*. Chapman & Hall/CRC, Boca Raton, FL, USA, 2nd edition.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4):430–450.

Grimes, P. W. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *Journal of Economic Education*, 33(1):15–30.

Grinblatt, M. and Keloharju, M. (2001). What makes investors trade? *The Journal of Finance*, LVI(2):589–616.

Grinblatt, M. and Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance*, LXIV(2):549–578.

Grossman, S. J. (1976). On the efficiency of competitive stock markets where traders have diverse information. *The Journal of Finance*, 31:573–585.

Grossman, S. J. (1981). An introduction to the theory of rational expectations under asymmetric information. *Review of Economic Studies*, XLVIII:541–559.

Gruca, T. S., Berg, J. E., and Cipriano, M. (2005). Consensus and difference of opinion in electronic prediction markets. *Electronic Markets*, 15(1):13–22.

Haigh, M. S. and List, J. A. (2005). Do professional traders exhibit myopic loss aversion? an experimental analysis. *The Journal of Finance*, IX(1):523–534.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hanson, R. (1999). Decision markets. *IEEE Intelligent Systems*, 14(3):16–19.

Hanson, R. (2007). Insider trading and prediction markets. George Mason University, Department of Economics.

Hastie, R. and Dawes, R. M. (2010). *Rational Choice in an Uncertain World: The Psychology of Judgement and Decision Making*. SAGE Publications, Inc., 2nd edition.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, 2nd edition.

Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review*, XXXV(4):519–530.

Healy, P. J., Linardi, S., Lowery, J. R., and Ledyard, J. O. (2010). Prediction markets: Alternative mechanisms for complex environments with few traders. *Management Science*, 56(11):1977–1996.

Heiberger, R. M. and Holland, B. (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS*. Springer Texts in Statistics, Springer.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons, Inc.

Hertwig, R., Herzog, S. M., Schooler, L. J., and Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5):1191–1206.

Hilton, D. (2001). The psychology of financial decision-making: Applications to trading, dealing, and investment analysis. *The Journal of Psychology and Financial Markets*, 2(1):37–53.

Ho, T.-H. and Chen, K.-Y. (2007). New product blockbusters: The magic and science of prediction markets. *California Management Review*, 50(1):144.

Hofmann, D. M. (2008). Dealing with the unexpected: Lessons for risk managers from the credit crisis. Technical report, Zurich Financial Services Group.

Hoggatt, A. (1959). An experimental business game. *Behavioral Science*, 4:192–203.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., and Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, 32:401–414.

Hubbard, D. W. (2009). *The Failure of Risk Management: Why It's Broken and How to Fix It*. John Wiley & Sons, Inc.

Huber, J. (2002). *Wahlbörsen: Preisbildung auf politischen Märkten zur Vorhersage von Wahlergebnissen*. Kovac, Hamburg.

Huizinga, J. (1987). An empirical investigation of the long run behavior of real exchange rates. *Carnegie-Rochester Conference Series on Public Policy*, 27:149–214.

Hung, M.-W. and Yu, H.-Y. (2006). A heterogeneous model of disposition effect. *Applied Economics*, 38:2147–2157.

Hutchinson, J. M. and Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, 69(2):97–124.

Jacquemin, A. and Berry, C. (1979). Entropy measure of diversification and corporate growth. *The Journal of Industrial Economics*, XXVII(4):359–369.

Janis, I. L. (1972). *Victims of groupthink.* Boston: Houghton Mifflin.

Jolliffe, I. T. (1986). *Principal Components Analysis.* Springer.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill Professional, 3rd edition.

Kagel, J. H. (1995). Auctions: A survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Markets.* Princeton University Press.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9):697–720.

Kahneman, D. and Frederick, S. (2002). *Heuristics and Biases*, chapter Representativeness Revisited:. Cambridge University Press.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206.

Kahneman, D. and Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1).

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3:430–454.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200.

Kaustia, M. (2010a). Disposition effect. In Baker, H. K. and Nofsinger, J. R., editors, *Behavioral Finance: Investors, Corporations, and Markets.* John Wiley & Sons, Inc.

Kaustia, M. (2010b). Prospect theory and the disposition effect. *Journal of Financial and Quantitative Analysis*, 45(3):791–812.

Kelly, W. (1981). A generalized interpretation of the herfindahl index. *Southern Economic Journal*, 48(1):50–57.

Kemeny, J. G. (1955). Fair bets and inductive probabilities. *The Journal of Symbolic Logic*, 20(3):263–273.

Kim, J.-O. and Mueller, C. W. (1978a). *Factor Analysis: Statistical Methods and Practical Issues*. series no. 07-014. Sage University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, California 91320, USA.

Kim, J.-O. and Mueller, C. W. (1978b). *Introduction to Factor Analysis*. series no. 07-013. Sage University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, California 91320, USA.

Kliger, D. and Kudryavtsev, A. (2008). Reference point formation by market investors. *Journal of Banking & Finance*, 32:1782–1794.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin Company.

Kumar, A. (2009). Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis*, 44(6):1375–1401.

Kumar, A. and Lim, S. S. (2008). How do decision frames influence the stock investment choices of individual investors? *Management Science*, 54(6):1052–1064.

Kuserk, G. J. and Locke, P. R. (1993). Scalper behavior in futures markets: An empirical examination. *The Journal of Futures Markets*, 13(4):409–431.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis os stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2):311–328.

Lichtenstein, S., Fischhoff, B., and Phillips, L. (1982). *Calibration of probabilities: The state of the art to 1980*. Cambridge University Press.

Likert, R. (1931). A technique for the measurement of attitudes. In *Archives of Psychology*. New York: Columbia University Press.

Litterman, R., editor (2003). *Modern Investment Management*. John Wiley & Sons, Inc.

Liu, Y. and Labrosse, F. (2002). Inverse validation for accurate range image registration with structured data. *IEEE*, 02:1051–1055.

Lo, A. W. and MacKinlay, A. C. (1988). Stock prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1:41–46.

Lo, A. W. and MacKinlay, A. C. (1989). The size and power of the variance ratio test in finite samples: A monte carlo investigation. *Journal of Econometrics*, 40:203–238.

Locke, P. R. and Mann, S. C. (2005). Professional trader discipline and trade disposition. *Journal of Financial Economics*, 76:401–444.

Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4:103–124.

Malkiel, B. G. (2007). *A Random Walk Down Wall Street*. W. W. Norton & Company, Inc., revised and updated edition edition.

Manaster, S. and Mann, S. C. (1996). Life in the pits: Competitive market making and inventory control. *The Review of Financial Studies*, 9(3):953–975.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Marshall, A. (1890). *Principles of Economics*. Macmillan and Co., Limited, 8th edition.

McCrae, R. R. and Costa, P. T. J. (1997). Personality trait structure as a human universal. *American Psychologist*, 52:509–516.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum, Mahwah, NJ.

McGraw, A. P., Mellers, B. A., and Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making*, 17:281–295.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.

Mill, J. S. (1848). *Principles of Political Economy*. D. Appleton and Company, New York.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2(3):255–273.

Mills, T. C. and Markellos, R. N. (2008). *The Econometric Modelling of Financial Time Series*. Cambridge University Press, 3rd edition.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, 55:1–22.

Moore, D. and Healy, P. (2007). Bayesian overconfidence. *Academy of Management Proceedings*, pages 1–6.

Mueller, D. C. (2003). *Public Choice III*. Cambridge University Press.

Mullainathan, S. and Thaler, R. H. (2000). Behavioral economics. Working Paper 00-27.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3).

Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295.

Nenadic, O. and Greenacre, M. (2007). Correspondence analysis in r, with two- and three-dimensional graphics. *Journal of Statistical Software*, 20(3):1–13.

Niederhoffer, V. and Osborne, M. F. M. (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, 61(316):897–916.

Nofsinger, J. R. (2011). *The Psychology of Investing*. Prentice Hall, 4th edition.

O'Connell, P. G. J. and Teo, M. (2009). Institutional investors, past performance, and dynamic loss aversion. *Journal of Financial and Quantitative Analysis*, 44:155–188.

Odean, T. (1998a). Are investors reluctant to realize their losses? *The Journal of Finance*, LIII:1775–1797.

Odean, T. (1998b). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance*, LIII(6):1887–1934.

Odean, T. (1999). Do investors trade too much? *The American Economic Review*, 89(5):1279–1298.

of Sponsoring Organizations of the Treadway Commission, C. (2004). Coso framework. Technical report, Committee of Sponsoring Organizations of the Treadway Commission.

of Sponsoring Organizations of the Treadway Commission, C. (2008). Internal control - integrated framework guidance on monitoring internal control systems. Technical report, Committee of Sponsoring Organizations of the Treadway Commission.

on Banking Supervision, B. C. (2006). International convergence of capital measurement and capital standards. Technical report, Basel Committee on Banking Supervision.

on Banking Supervision, B. C. (2010, revised 2011). Basel iii: A global regulatory framework for more resilient banks and banking systems. Technical report, Basel Committee on Banking Supervision.

Pennock, S., Lawrence, C., Giles, L., and Nielsen, F. (2001). The real power of artificial markets. *Science*, 291:987–988.

Persky, J. (1995). The ethology of homo economicus. *The Journal of Economic Perspectives*, 9:221–231.

Phillips, P. and Perron, P. (1988). Testing for unit roots in time series regression. *Biometrika*, 75:335–346.

Plott, C. and Sunder, S. (1982). Efficiency of experimental security markets with insider information: An application of rational expectations models. *Journal of Political Economy*, 90:663–698.

Plott, C. and Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56:1085–1118.

Plott, C. R. (2000). Markets as information gathering tools. *Southern Economic Journal*, 67(1):1–15.

Plous, S. (1993). *The Psychology of Judgement and Decision Making*. McGraw-Hill, Inc.

Prelec, D. (2000). Compound invariant weighting functions in prospect theory. In Kahneman, D. and Tversky, A., editors, *Choices, Values, and Frames*. Cambridge University Press.

Presson, P. and Benassi, V. (1996). Illusion of control: A meta-analytic review. *Journal of Social Behavior and Personality*, 11(3):493–510.

Project Management Institute, I. (2004). *A Guide to the Project Management Body of Knowledge*. Project Management Institute, Inc., 3rd edition.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R., editor, *The Foundations of Mathematics and other Logical Essays*, chapter VII, pages 156–198. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company, 1931.

Ray, R. (2006). Prediction markets and the financial "wisdom of crowds". *The Journal of Behavioral Finance*, 7(1):2–4.

Reilly, F. and Brown, K. (2006). *Investment Analysis and Portfolio Management*. Thomson South-Western, 8th edition edition.

Revelle, W. (2011). psych. *Procedures for Psychological, Psychometric, and Personality Research*, 1.1.12.

Rhode, P. W. and Strumpf, K. S. (2004). Historical presidential betting markets. *The Journal of Economic Perspectives*, 18(2):127.

Rosenbloom, E. S. and Notz, W. (2006). Statistical tests of real-money versus play-money prediction markets. *Electronic Markets*, 16(1):63–69.

Roth, A. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212.

Russo, J. E. and Schoemaker, P. J. H. (1989). *Decision Traps: Ten barriers to brilliant decision-making an dhow to overcome them*. New York: Doubleday.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, New Series*, 5(17):61–71.

Samuelson, P. A. (1957). Intertemporal competitive equilibrium: An empirical study of speculation. In *The Theory of Speculation*, volume 79, chapter 8, pages 181–221. Springer: Weltwirtschaftliches Archiv.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–50.

Sander, D. and Scherer, K. R. (2009). *The Oxford companion to emotion and the affective sciences*. Oxford University Press.

Santayana, G. (1905). *The Life of Reason; or the Phases of Human Progress*, volume I. Charles Scribner's Sons.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons, Inc.

Scheinkman, J. and Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6):1183–1219.

Scott, M. (2008). Euro 2008 kicks economies up a notch. *BusinessWeek*.

Selten, R. (2001). What is bounded rationality? In Gigerenzer, G. and Selten, R., editors, *Bounded Rationality: The Adaptive Toolbox*. The MIT Press.

Servan-Schreiber, E., Wolfers, J., Pennock, D., and Galebach, B. (2004). Prediction markets: Does money matter? *Electronic Markets*, 14(3):243–251.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Sheather, S. J. (2009). *A Modern Approach to Regression with R*. Springer Texts in Statistics, Springer.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, XI(3).

Sherden, W. A. (1998). *The Fortune Sellers: The Big Business of Buying and Selling Predictions*. John Wiley & Sons, Inc.

Siegel, A. (2008). Taking the pulse of your company: Charting risk with prediction markets. *Risk Management*, 55(4):52.

Siegel, S. and Fouraker, L. (1960). *Bargaining and Group Decision Making*. New York: McGraw-Hill.

Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Simon, H. A. (1957). *Models of Man*. Wiley.

Singal, V. and Xu, Z. (2011). Selling winners, holding losers: Effect of fund flows and survival of disposition-prone mutual funds. *Journal of Banking & Finance*, 35:2704–2718.

Siwar, E. (2011). The impact of overconfidence bias and disposition effect on the volume of transaction and the volatility of the french stock market. *Journal of Applied Economic Sciences*, VI(1/15):61–83.

Skousen, M. (2009). *The Making of Modern Economics*. M. E. Sharpe, 2nd edition.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119:3–22.

Slovic, P. (1969). Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology*, 53(4):255–263.

Slovic, P. (1972). Psychological study of human judgment: Implications for investment decision making. *The Journal of Finance*, 27(4):779–799.

Slovic, P. and Lichtenstein, S. (1971). Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organization Behavior and Human Performance*, 6:649–744.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford University Press (1960).

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70:111–137.

Smith, V. L. (1982). Microeconomic systems as an experimental science . *The American Economic Review*, 72:923–955.

Smith, V. L. (2008). *Rationality in Economics*. Cambridge University Press.

Stanovich, K. E. and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23:645–665.

Stasser, G. and Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry*, 14(3 & 4):304–313.

Stix, G. (2008). When markets beat the polls. *Scientific American*, 298(3).

Sugiura, M. and Sasaki, M. (2011). Nuclear expert: Hubris led to disaster. online.

Summers, B. and Duxbury, D. (2007). Unraveling the disposition effect: The role of prospect theory and emotions. Available at SSRN: http://ssrn.com/abstract=1026915 or doi:10.2139/ssrn.1026915.

Sunder, S. (1995). Experimental asset markets: A survey. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics*. Princeton University Press.

Sunstein, C. R. (2006). *Infotopia*. Oxford University Press.

Surowiecki, J. (2004). *The Wisdom of Crowds*. Anchor Books, A Division of Random House, Inc.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47:143–148.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1:39–60.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3):199–214.

Thaler, R. (1987a). Anomalies: Seasonal movements in security prices ii: Weekend, holiday, turn of the month, and intraday effects. *Journal of Economic Perspectives*, 1(1):169–177.

Thaler, R. and Johnson, E. (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.

Thaler, R. and Ziemba, W. T. (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, 2(2):161–174.

Thaler, R. H. (1987b). Anomalies: The january effect. *Economic Perspectives*, 1(1):197–201.

Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, 89(21):392–406.

Thurston, L. L. (1931). The indifference function. *The Journal of Social Psychology*, 2:139–167.

Todd, P. M. and Gigerenzer, G. (2001). Putting naturalistic decision making into the adaptive toolbox. *Journal of Behavioral Decision Making*, 14:353–384.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323.

Vastag, B. and Mufson, S. (2011). Japanese scientist: Fukushima meltdown occurred within hours of quake. online.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wakker, P. P. (2010). *Prospect Theory for Risk and Ambiguity*. Cambridge University Press.

Walras, L. (1874). *Éléments d'économie politique pure, ou théorie de la richesse sociale*. (Elements of Pure Economics, or the theory of social wealth, transl. W. Jaffé, 1899).

Weber, M. and Camerer, C. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33:167–184.

Whyte, G. (1989). Groupthink reconsidered. *Academy of Management Review*, 14(1):40–56.

Wikipedia (2011). http://en.wikipedia.org/wiki/list_of_cognitive_biases. retrieved 27.09.2011.

Wilson, R. (1992). Strategic analysis of auctions. In Aumann, R. J. and Hart, S., editors, *Handbook of game theory with economic applications*, volume 1. Amsterdam: Elsevier Science Publishers.

Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining*. Morgan Kaufmann Publishers, 3rd edition.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *The Journal of Economic Perspectives*, 18(2):107–126.

Wolfers, J. and Zitzewitz, E. (2006). Prediction markets in theory and practice. Working Paper 12083, NBER Working Paper Series, National Bureau of Economic Research.

Wong, A. S., Carducci, B. J., and White, A. J. (2006). Asset disposition effect: The impact of price patterns and selected personal characteristics. *Journal of Asset Management*, 7(3/4):291–300.

www.biz.uiowa.edu/iem. Iowa electronic markets. online. retrieved 23.9.2009.

Yazdipour, R. and Howard, J. A. (2010). Behavioral finance: Application and pedagogy in business education and training. In Baker, H. K. and Nofsinger, J. R., editors, *Behavioral Finance: Investors, Corporations, and Markets*. John Wiley & Sons, Inc.

Curriculum Vitae

Donald Baillie — Curriculum Vitae

Personal Data

Name:	Donald Baillie
Date of birth:	7. April 1968
Nationality:	New Zealand
Address:	Porzellangasse 50/15 A-1090 Vienna, Austria
Telephone:	+43-676 30 25 240
e-mail:	donald.baillie@anzacfinance.at

Education

- Vienna University of Technology, Austria
  Doctoral Studies in Business Administration, 2009 - present
- Vienna Economics University, Vienna, Austria
  Master of Economics and Business Administration, 1994

Other Qualifications

- Awarded the title of "Fachhochschul-Dozent" (FH-Dozent), University of Applied Science Vienna, 2009
- Certified Basel II Consultant, Austrian Chamber of Commerce, 2006
- GARP: Financial Risk Manager Certification (FRM), 2005
- Certified Trader and Market Maker for Securities Trading
  Vienna Stock Exchange, April 2002
- Certified Trader and Market Maker for Futures and Options Trading
  Vienna Stock Exchange, April 2002

Work Experience

Oct.2011 – present	**Managing Director** **ifb Austria GmbH**
Responsibilities:	General Manager of the Austrian legal entity of an international consulting company, responsible for all personnel, business and market operations. Responsible within ifb International for risk management topics and consulting personnel associated with risk management consulting.
Jan.2005 – Dec.2011	**General Manager** **ANZAC Finance Solutions GmbH**
Responsibilities:	Project acquisition and management for Treasury and Risk Management projects, development of the firm, cooperation with business partners.
Jan. 2003 - present May 2006 - present*	**Lecturer and Member of the University Council** **University for Applied Science (Fachhochschule für Finanz-, Rechnungs- und Steuerwesen) and Webster University Vienna***
Responsibilities:	Lecturing at the faculties of Finance. Lecture topics: Corporate Finance, Treasury Management, Risk Management. In addition to tasks as lecturer, supervisor for Masters' Theses in finance. * Lecture topics: Undergraduate Investments, Graduate Investments, Graduate Derivatives

Aug. 2002 – Dec.2004	**General Manager** **SPIRIT Business & Finance Solutions GmbH**
Responsibilities:	Responsible for the acquisition and operation of all consulting and project management activities as well as the development of the firm in the market and internally. Two focus areas: strategic projects in Treasury (proprietary trading and sales) & Risk Management (market risk, operational risk, Basel II) and software development for these areas.
Jan. 2000 – Jun. 2002	**Senior Manager/Principal Consultant** **Ernst & Young Management Consulting Austria (now Cap Gemini)**
Responsibilities:	Responsible (personnel and budget) for a group of 10 people focused on treasury and risk management consulting. Co-headed and managed all operations of the subsidiary performing the development of the market risk management product *RiskVantage*, adding modules for liquidity management and fund performance measurement. Responsible for the acquisition of new clients, new projects at existing clients, project management of all implementations of the software at clients' sites, as well as ensuring on-going support activities for these clients. Performed account management for additional selected clients in the banking sector, as well as acquisition for all Austrian accounts involving treasury and risk management topics. Led numerous projects in banking, IT, insurance and industry. 2001: Co-led the project to search for a buyer for the risk management subsidiary firm CGE&Y Financial Advisory Services, performed due diligence, all legal and commercial negotiations and subsequently sold the firm.
Jan.1999 – Dec.1999	**Manager** **Ernst & Young Management Consulting Austria**
Responsibilities:	Led numerous projects in the business lines banking, insurance and industry concerning treasury, risk management, asset management, package-enabled reengineering, IT implementations and general banking business process reengineering. Subject matter expert for wholesale banking, treasury, risk management and financial theory. Member of the Ernst & Young International Risk Management Network. Responsible for building up a group of consultants in the wholesale banking business and risk management. Led the development of the software product *RiskVantage* for market risk management and CAD reporting, and managed the projects implementing the software at clients' sites.
Nov.1996 – Dec.1998	**Market Risk Management** **Creditanstalt AG bank, Vienna, Austria**
Responsibilities:	Performed day-to-day market risk management for global bank operations. Developed and implemented the first Austrian internal model based on VaR for risk management and for external CAD reporting, subsequently approved by the Central Bank. Implemented the internal model in selected foreign branches, and supported them on a daily basis. Performed daily backtesting and regular stresstesting as well as daily reporting, limit control and monthly ALCO reporting.
Sep.1994 – Oct. 1996	**Treasury Organisation** **Creditanstalt AG bank, Vienna, Austria**
Responsibilities:	Performed organisation and support for Treasury. Led and participated in projects focusing on both IT systems implementation as well as development and proof of financial models for Money Market and FX derivatives (p.e. implementation of Opus front- and back office system for interest rate derivatives). Conducted tests for financial models and systems implemented in Treasury for head office and all worldwide branches (p.e. for WallStreet Systems and Opus). Maintained constant contact with global Treasury staff and performed coordination and prioritisation of day-to-day

Donald Baillie
A-1090 Vienna, Porzellangasse 50/15

	issues concerning financial models and systems. Produced a bi-weekly bankwide Treasury newsletter.
Dec. 1993 – Jan. 1994	**Software Developer** **CogiData EDV-Service Fürst KG, Vienna, Austria**
Responsibilities:	Participated in a project developing in COBOL a warehousing software package for a large oil company. Performed the development of the modules responsible for the database management of the software.
Oct. 1991 – Jun. 1992	**Trainer** **Computer Schule Millergasse, Vienna, Austria**
Responsibilities:	Conducted beginners' and advanced training courses in MS-DOS, MS Windows and all Windows software, organised curriculum and day-to-day operation.
Jul. 1990 – Apr. 1991	**Software Developer and Support Technician** **CogiData EDV-Service Fürst KG, Vienna, Austria**
Responsibilities:	Developed software using COBOL, C, Pascal. Performed hardware and software support for PC's, HP-3000 and HP-9000 systems running DOS, Unix, MS-Windows and HP-MPE V. Installed and maintained software and hardware internally and at customer sites.

Publications

- Paper co-author: "Overconfidence in Sports Betting", International Congress of Psychology, forthcoming
- Article author: "Trading with Weather Derivatives – the Situation in Austria", CFO Aktuell, March 2010 (in German)
- Article author: "Weather Derivatives – an innovative Financial Product", CFO Aktuell, February 2010 (in German)
- Book author: "Doing Business with the Weather" (in German), Grelldenk Verlag, 2009
- Book co-author: „The 100 Most Important Topics in Business Administration for Practitioners" (in German), Linde International, 2005
- Report co-author: „Banking Risk Disclosure in Annual Reports", Ernst&Young, 2000
- Article author :"Risk Management", Slovakian Banking Journal, 1999

Languages

- Mother Tongue: Bilingual in English and German
- Conversational French