Forecasting Economic Indices

Design, Performance, and Learning in Prediction Markets

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Wi.-Ing. Florian Teschner

Tag der mündlichen Prüfung: 27.07.2012
Referent: Prof. Dr. Christof Weinhardt
Korreferent: Prof. Dr. Orestis Terzidis

2012 Karlsruhe

Acknowledgements

I would like to express my gratitude and thanks to my advisor Prof. Dr. Christof Weinhardt for his support, trust, and advice throughout the preparation of this dissertation. He always knew the right time to encourage me and the right time to challenge me. His enthusiasm and visionary thinking have been an inspiration for my research.

I would also like to thank Prof. Dr. Orestis Terzidis, Prof. Dr. Ryan Riordan, and Prof. Dr. Martin Klarmann, who have served on my thesis committee, for spending much time on reading this thesis and providing insightful commentary on my work.

My sincere thanks go to my colleagues at the Institute of Information Systems and Management (IISM). It has been a privilege to be a member of this inspiring team. In particular I would like to thank my colleagues Tobias Kranz, Dr. Stephan Stathel, Christoph Flath, Dr. Martin Wagener, Dr. Andreas Storkenmaier, and Sarah Zhang. Discussions with or suggestions from them at different stages were very rewarding.

I am also very grateful to Dr. David Rothschild and Dr. David Pennock of Yahoo! Research. They gave me the opportunity to shed light on new aspects of my work and included in me in their fantastic research group. I gratefully acknowledge that the research stay was supported by a scholarship of the Karlsruhe House of Young Scientists (KHYS).

Last, but by no means least, I thank Isabel for her understanding and patience.

Finally, but most importantly, I would like to thank my family, especially my parents Anke and Micha for their support and caring encouragement.

Contents

Chapter 1

Introduction & Motivation

1.1 Motivation

A wide range of policy decisions is made on the basis of economic forecasts. Inaccurate or delayed predictions can result in substantial costs for companies and citizens. However, it is an established fact that traditional economic forecast models lack accuracy (McNees, 1992; Schuh, 2001; Osterloh, 2008). The recent financial crisis exemplified the failure of economic forecasting. Weeks after Lehman Brothers filed for bankruptcy protection, the consensus still predicted a 2 % rise in German GDP for 2009, but it dropped by 4.5 %. Most often contemporary forecast methods mix expert knowledge and extrapolation using historical data. They are often unable to capture economic shocks (Clements and Hendry, 2002). In Germany, forecasts are provided by numerous institutions and released on a periodical basis. Additionally, these forecasts vary in time horizon and definition. Thus decision makers may find it difficult to aggregate various forecasts and derive a robust appraisal. However, economic uncertainty is seen as a major challenge for management and policy-making in the 21st century (Greenspan, 2004; Nitin and Stewart, 2006). To conclude, there is no central mechanism to continuously combine individual forecasts. Combining and eliciting this information and forecasts would mean to identify experts, to motivate their participation, and to determine how to aggregate different opinions.

With the growth of the Internet, markets that trade predictions about future events have emerged as a promising alternative forecasting tool. In these markets, participants trade contracts whose payoff depends on the outcome of uncertain future events. For example, a market contract might reward a dollar if a particular presidential candidate is elected. An individual who thinks the candidate has a 65 % chance of being elected should be willing to pay up to 65 cents for such a contract. Market participants form expectations about the outcome of an event. Comparable to financial markets, they buy if they find that prices underestimate probability of the event in question and they sell a stock if they find that prices overestimate the probability of an event. Also called prediction markets, these markets thereby aggregate information in the same way as a stock market. Market prices represent the participants' aggregated expectations and can be interpreted as the market forecast (Gjerstad and Hall, 2005).

Prediction markets have a long track of successful application in a wide area ranging from political elections (Berg et al., 2008) to sport events (Luckner et al., 2008) sometimes outperforming established forecast methods (Spann and Skiera, 2004).

As a forecasting method prediction markets offer many advantages. They provide the incentives for traders to truthfully disclose their information and an algorithm to weight opinions (Arrow et al., 2008). Compared with statistical forecasting methods, these markets can incorporate real-time information. As prediction market prices are updated immediately when traders incorporate their expectation in prices, they provide continuously and timely updated forecasts. Compared with eliciting expert opinions, prediction markets eliminate the effort of identifying experts and motivate their participation. In most cases they allow anonymous participation, which may increase the likelihood of nonconformists to participate and reveal information and they do not need to deal with conflicting opinions.

Prediction markets communicate feedback on two levels. Market prices provide the first feedback. Traders recognize prices as the current aggregated forecast and incorporate this information in their own beliefs. Secondly, after contracts are paid out, participants realize their gains and losses. This provides traders with feedback about their own trading performance. Moreover, as good traders increase their portfolio value over time, they gain more weight over market run-time. To sum up, evidence so far suggests that prediction markets provide considerable advantages in terms of continuous forecasting, participation and information revelation.

We thus setup a prediction market called Economic Indicator Exchange (EIX)[1] to forecast macroeconomic variables. The EIX play-money prediction market is specifically designed to continuously and repeatedly forecast economic indicators such as GDP, inflation, Ifo index, investments, export and unemployment figures in Germany. The goal is to forecast these indicators over long time horizons and continuously aggregate economic information. In order to build a sustainable market with long-term participation, the EIX was launched in cooperation with the leading German economic newspaper "Handelsblatt". The cooperation aims at reaching a wide and well-informed audience interested in financial markets and economic development. The market is publicly available over the Internet and readers are invited to join.

The advantages of this research setting are at least threefold. To begin with, market forecasts can be benchmarked against well-established forecasting methods. Secondly, from an individual perspective, market participants interact in a repeated decision-making environment closely reassembling decision-making in financial markets. As the outcome of events in prediction markets is finally known, we can ex-post measure the participants' trading performance. Finally, with over 1,300 participants the EIX-market offers a good platform to run web-based experiments.

1.2 Research Outline & Questions

This thesis aims to analyze the predictive power of a prediction market for macroeconomic forecasts. Within the scope of this work, we[2] address the following research questions:

[1] www.eix-market.de or eix.handelsblatt.com

[2] The *We* refers to both; the readers of this work and my co-authors. I would like to particularly thank Christof Weinhardt, Ryan Riordan, Stefan Stathel, Athanasios Mazarakis, Tobias Kranz, Maximilian Coblenz, Christoph Schiller, Stefan Rehm.

RQ1. How to quantify prediction market performance and quality? Over the last couple of years, interest in prediction markets as a forecasting tool has continuously increased in the scientific world and in industry. Even though there are numerous empirical prediction market studies, no statistical or quantitative meta-analysis were conducted to analyze prediction market accuracy for different problem types or application areas. One reason is that there is no unified, stringent methodology used to analyze prediction market data. Moreover, there is no general understanding what defines prediction market performance. In finance the ambiguous term of market quality mostly refers to liquidity measures (e.g. spreads) or information measures such as permanent price impact. Prediction markets are often quantified solely through the correlation between predictions (derived from prices) and event outcomes. However, judging prediction market quality on the basis of forecast errors hinders the understanding of the underlying factors. Therefore there is a need to summarize and evaluate market performance measures for prediction markets.

RQ2. How to design a market to forecast macroeconomic indices? Macroeconomic forecasts are used extensively in industry and government even though the historical accuracy and reliability is disputed. In Germany forecasts are provided by numerous governmental or industry sponsored institutions and released on periodical basis. Thus, business and governmental decision makers might find it difficult to aggregate the various forecasts and form a confident belief. As prediction markets have a track of successful application (e.g., Berg et al., 2008; Luckner et al., 2008), it seems promising to design a prediction market to forecast economic indices. However, there is a lack of understanding how to design and run a prediction market for macroeconomic forecasting.

RQ3. What is the effect of short-selling on market quality and forecast performance? Shortly after the beginning of the global financial crisis, several stock exchange regulators banned short selling on financial markets (SEC, 2008). They argued that this was necessary to maintain properly operating markets for securities trading and to ensure the stability of the financial system. An ongoing debate in finance is whether short selling has positive or negative effects on market quality. Introducing short selling in a prediction market setting facilitates the analysis of its impact on market quality and forecast performance.

RQ4. How do incentives and feedback mechanisms affect participation in prediction markets? An arising question is how to design incentive schemes and feedback mechanisms to motivate participants in online communities to contribute. Subsequently this leads to the question of how participants can be motivated to contribute and share their information for longer time horizons. In public goods projects participation feedback has been found to increase participants' contributions (Cheshire, 2007). In order to motivate participants intrinsically, we employ these feedback mechanisms in a prediction market setting and try to find out (i) which of the feedback types works best at motivating contributions and (ii) do the additional contributions improve the forecasts?

RQ5. How do trading interfaces affect trading behavior and performance? Market interfaces tend to be complex and non intuitive. Users not familiar with such interfaces find it hard to interact. Recently, researchers have identified the need to merge interface and market design to address this issue (e.g. Seuken et al., 2010). The goal is to hide the market or to design the interface in a more intuitive way and thus simplify the market interface. One way to accomplish that is to simplify the market interface but maintain economic efficiency. Hence, there is a need to understand which information elements support and which elements hinder the individual trading process. Moreover, as individuals have different informational needs and vary in experience, it seems fruitful to develop customized market interfaces.

1.3 Structure of the Thesis

This thesis is organized following the order of the before mentioned research questions. Chapter 2 reviews related work in market engineering and prediction markets. It additionally provides a prediction market quality framework by summarizing financial market quality measures and measures for forecast performance. Chapter 3 details the Economic Indicator eXchange (EIX), a prediction market for macroeconomic forecasts. It discusses market and interface design choices as well as the course of the two-year field experiment. Chapter 4 describes the market by applying the before mentioned market quality framework. We evaluate the market from both a market engineering as well as a forecasting performance perspective.[3] Chapter 5 analyzes the effect of introducing short selling on market quality.[4] The subsequent chapter discusses the effect of feedback and incentives on learning in markets.[5] Addressing the intersectional topic between market and interface design, Chapter 7 highlights the importance of consciously designing market interfaces. Firstly, the chapter examines the effect of different interface elements on trader behavior and performance. Secondly, it compares trader performance between a simplified and a standard market interface.[6] Finally, Chapter 8 summarizes the key contributions of the work at hand and outlines promising topics for future research.

[3]Parts of this chapter are joint work with Stephan Stathel and Christof Weinhardt and have been presented at the Hawaii International Conference on System Sciences 2011 (Teschner et al., 2011) and the Americas Conference on Information Systems 2012 (Teschner and Weinhardt, 2012a)

[4]The results of this section are published in the Journal of Prediction Markets (joint work with Maximilian Coblenz and Christof Weinhardt (Teschner et al., 2011))

[5]The findings have been presented at the International Conference on Information Systems 2011 (joint work with Athanasios Mazarakis, Ryan Riordan and Christof Weinhardt (Teschner et al., 2011)) and the European Conference on Information Systems 2012 (Teschner and Weinhardt, 2012b).

[6]Selected parts of this chapter have been presented at the Second Conference on Auctions, Market Mechanisms and Their Applications (Teschner and Weinhardt, 2011) and the Doctoral Consortium 2011, Wirtschaftsinformatik (Teschner, 2011)

Chapter 2

Related Work & a Prediction Market Framework

IN 2005 James Surowiecki coined the term *the wisdom of crowds* by describing how groups of people solve, under certain conditions, complex problems far better than single individuals (Surowiecki, 2004). There are various ways to utilize the wisdom of crowds such as using wikis, reputation systems, or polling mechanisms. Another way to aggregate dispersed information is by setting up a so called prediction market [1]. Over the last couple of years, interest in prediction markets as a forecasting method has continuously increased in the scientific world and in industry (Tziralis and Tatsiopoulos, 2007). In general, markets fulfill at least two roles: allocate resources and aggregate information about the value of resources (Hayek, 1945). These two roles are best illustrated by a short example. 90% of all US oranges used in frozen concentrated orange juice are grown in central Florida. Clearly this makes the weather in central Florida of critical importance to the future supply of orange juice. Hence traders of *orange-concentrate* futures have a good incentive to sell if they belief the weather to be good for the orange harvest, and vice versa. The market price for these futures is the aggregate belief off all individual forecasts. Roll (1984) shows that *orange-concentrate* prices indicate weather extremes more precisely than the US national weather service. Thereby markets provide incentives for information revelation

[1] In literature various names are used such as idea futures, information market, virtual stock market, artificial and prediction market with the term *prediction market* the most prevailing.

and acts as a mechanism for aggregating information.
This work is focused on the informational perspective of markets. Moreover, it is concerned with the question of how to design and engineer markets to improve their ability to efficiently aggregate information.

2.1 Market Engineering

An important question in this context is how to design and engineer (electronic) markets so that they become successful and can actually *deliver* the transparency and market efficiency they promise. Following (Weinhardt and Gimpel, 2006, p. 6) we define a market as *"a set of humanly devised rules that structure the interaction and exchange of information by self-interested participants in order to carry out exchange transactions at a relatively low cost."* In order to consciously design markets delivering the desired outcome, Weinhardt et al. (2003); Neumann (2004); Weinhardt et al. (2006) propose the *Market Engineering Framework* (See Figure 2.1, left side). The goal is to define a structured, systematic, and theoretically founded procedure of designing, implementing, evaluating, and introducing electronic market platforms.
The objective of a market engineer is to achieve a desired market outcome or performance. To do so, she can design the transaction object as well as the market structure. The market structure comprises the market microstructure, the (IT-) infrastructure, and the business structure. These designed elements, the transaction object and the market structure, have only indirect effect on the market outcome. The link from the structure to the outcome lies in the behavior of market participants. Usually, market engineers employ a variety of methodologies to asses the impact of specific market structures on the participants' behavior and thus the outcome. These methods include theoretical modeling (game theory, auction theory, mechanism design) and empirical research such as lab and field experiments. Until recently there was no common understanding how to empirically analyze markets and measure their quality. For financial markets Zhang et al. (2011) present a methodology framework to fill this gap.
To begin with, the socio-economic and legal environment comprises elements

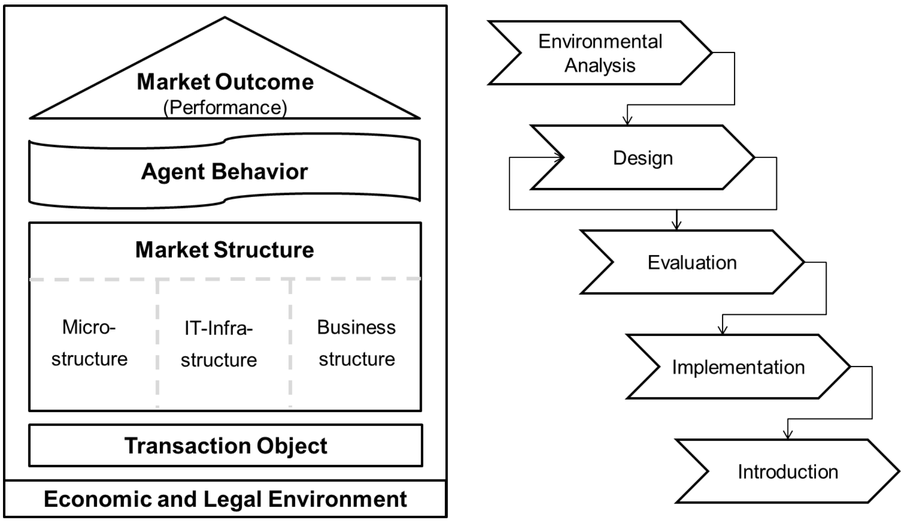

Figure 2.1: Market Engineering Framework (left) and Process (right) (Weinhardt et al., 2006)

which cannot directly be influenced. Examples for these elements are the participants' cultural background and norms, their preferences, and the applicable laws. One part that directly influences and mediates user behavior and which is outside the framework's scope is the market's graphical user interface (GUI). Just recently Seuken et al. (2011) have highlighted its importance for the success of markets.

Another part of the overall approach is a structured market engineering process (Weinhardt et al., 2007), which defines a best practice process to develop and implement markets (See Figure 2.1, right side). The process incorporates five stages; In the environmental analysis the requirements of the new electronic market are deduced. The second stage of the process comprises the design of the market with the simultaneous consideration of the market structure. Having designed the market, it is tested upon its technical and economic properties. This is followed by the implementation stage in which the design is realized and implemented as a software system. Finally, the introduction of the electronic market initiates its operation cycle (Neumann, 2004).

In order to keep up with recent developments in software engineering (rapid, test-driven development), (Block, 2010, p. 99) adapts the process to an agile market engineering process. *"In its core, the agile market engineering process model* *builds on short, incremental market development cycles and frequent user feedback in order to develop and to continuously refine and improve the electronic market platform."* Thus far it remains unclear how rapid continuous market improvements can be aligned with a rigorous scientific method.

As presented a wide range of design considerations have to be taken into account to develop a market. Each market has different objectives and requirements and hence needs to be carefully engineered. In the next section, markets designed to aggregate dispersed information for forecasting purposes, so called *Prediction Markets*, are introduced. Thereafter, design parameters for prediction markets will be discussed.

2.2 Prediction Markets

Prediction markets have proven to successfully forecast events in a wide range of applications. They facilitate and support decision making through aggregating expectations about events (Hahn and Tetlock, 2006). The roots of their predictive power are twofold: the market provides the incentives for traders to truthfully disclose their information and an algorithm to weight opinions (Arrow et al., 2008).

The most basic trading mechanism for prediction markets is based on a continuous double auction for one stock which represents the outcome of an event. The stock will pay a predefined value (e.g. 100) if an event has the predicted outcome and else the stock will be worthless. Market participants hold beliefs about the likelihood of an event. Comparable to financial markets, they buy if they find that prices underestimate the event in question and they sell a stock if they find that prices overestimate the probability of an event. A famous example is the Iowa political stock market (PSM) which predicts the outcome of U.S. presidential elections. The Iowa PSM features contracts that represent one nominee each. Market participants buy and sell nominee contracts depending on their assessment of the U.S. presidential election outcome. The U.S. presidential elections are well suited for a prediction market. In the final pre-election period only two candidates have a chance of winning the election which gives the market two

complementary assets. Also, only one of the nominees will win and the other one will lose. The first stock pays 100 if the second is judged at 0 and vice versa. Which means a stock pays 100 if the corresponding nominee wins an election. Usually this market design offers the possibility to buy and sell bundles of both stocks for 100 which implies that in a frictionless world with rational traders both stock prices always sum up to 100 (Berg et al., 2000).
The above described form of representing a single event with two complementary stocks has been the default since its proposal. In fact the concept has been so successful that it was adapted for events with more than two outcomes. For example, Luckner and Weinhardt (2008) design a market to predict the outcome of the FIFA soccer world cup 2006 where the value of all 32 stocks combined is predefined. All traded stock prices are dependent as there is by definition only one world champion.

2.2.1 Prediction Market Framework

Even though there are numerous empirical prediction market studies; no meta-analysis has been conducted that analyzed prediction market accuracy for various types of problems and fields of application. One reason for that is that there is no unified, stringent methodology used to analyze prediction market data. Although there are a some published guidelines (Luckner, 2008; Sripawatakul and Sutivong, 2010; Plott and Chen, 2002) on how to design and implement a prediction market, no work has yet been done on an evaluation methodology. Moreover, there is no general understanding what defines prediction market quality.Hence, we present an array of market quality measures which can enrich the prediction market methodology.
Due to a missing precise definition of market quality, we provide a methodology framework that puts several prominent measures of market quality into the context of internal design factors. Its structure and the outline of the following section is inspired by Zhang et al. (2011). The proposed framework for prediction market quality (Figure 2.2) consists of three main parts which will be discussed separately in the following subsections:

1. The forecasting goal.
2. The internal market design fundamentals such as the market microstructure, the incentives and the technology.
3. And quantifiable measures of market quality in terms of activity, liquidity, information and forecast performance.

The forecasting goal and description of the prediction issue influence market quality indirectly. In contrast, the internal market design concerns how the market is designed, incentivized, and implemented. Finally in order to implement and run the market, the operators can use a wide variety of IS-technologies. The technology used affects the trading system performance as well asthe intuitiveness of the user interface.

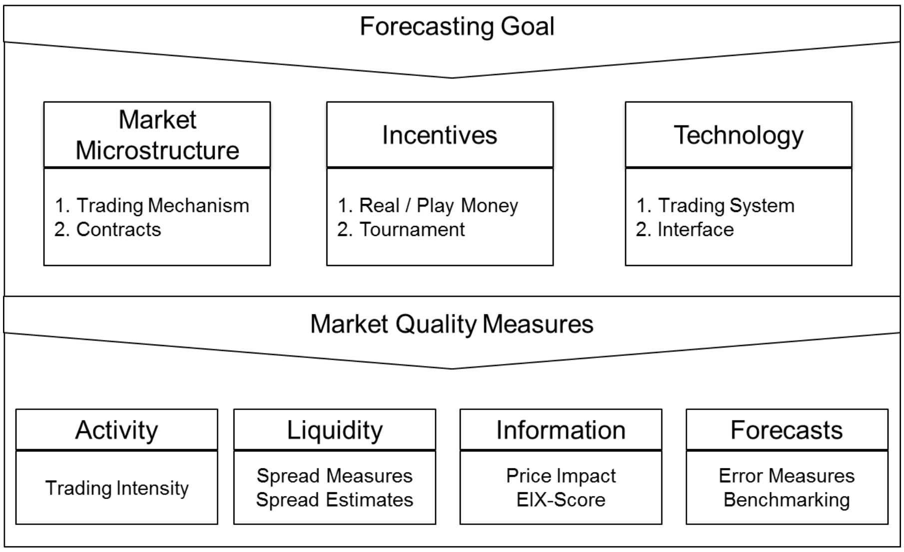

Figure 2.2: Prediction market quality framework, following Zhang et al. (2011)

Forecasting Goal

We see the forecasting goal as externally given, for example the prediction of product sales in the next month. Thus it is not under the explicit influence of the market designer. This means that an event is unambiguously the basis for the cash dividend of the respective shares of stock. However, the determination of the cash dividend needs to be clearly communicated to the market participants. Additionally, the participants need to have some knowledge about the event outcome, otherwise stock prices are randomly set (Forsythe et al., 1992). The most common prediction deals with the occurrence or non-occurrence of a particular event. This is often represented as a binary outcome. This is within the design considerations of the market designer. A specific outcome can usually be represented in various ways. The Iowa presidential market is a prime example. The outcome is the president to be elected. One design could feature the vote-share a nominee achieves. Another could be whether one nominee wins the election. Essentially the information gathered is very similar. However the vote-share design is a linear representation of a binary outcome (Berg and Rietz, 2006). Other possible prediction issues corresponding to specified events could be the prediction of an absolute number, such as the unemployment number in a certain month or the prediction of a relative number, for example the market share in a particular period. The appropriate formulation for the payoff function in both cases could be a linear transformation. Empirical results from political stock markets indicate that prediction markets work for different payoff functions as long as they are comprehensible for participants (Forsythe et al., 1999).

Market Design

The key design elements comprise the contract specification, the choice and details of the trading mechanism, the incentives provided, and the trading technology.

The central aspect of any trading platform is how buyers and sellers are matched. The most popular method is the continuous double auction (CDA). In a CDA, information is continuously updated as traders use news and events to revise their decisions and make a profit by acting accordingly. In this structure, traders place buy (bid) or sell (ask) orders for a contract with a specific volume, time validity and price. When the prices match or the ask price is lower than the bid price, a trade occurs. This continues until all matching orders have been fulfilled and a gap (spread) exists between the bid and ask prices. As the mechanism only matches willing traders, it can be implemented as a zero-sum game meaning that the market operator cannot lose money. As a consequence this mechanism

is especially useful for real-money exchanges. Alternative mechanisms are call auctions, (dynamic) pari-mutual markets (Pennock, 2004) and market scoring rules (Hanson, 2003). A discussion of the mechanisms' differences can be found in (Christiansen, 2007; Luckner, 2008).
Three common contract types are winner-takes-all, index and conditional contracts. In the winner-takes-all or digital markets only one contract is paid out whereas all others are valued zero in the end.Usually the market offers a portfolio trade where one share of each contract can be bought for 100. Traders have the opportunity to gain money by exploring arbitrage opportunities in two ways. First by buying a portfolio and selling the contracts separately. If the sum of the single contracts exceeds 100, participants can buy all contracts for 100 and sell them individually. Second if the sum of all contracts' prices is below 100, traders can profit by buying individual contracts and selling the portfolio bundle for 100. In index markets, shares in the market are paid out according to a certain percentage. Referring to the presidential markets if the market's question is "What vote share will a candidate achieve?" The contract is worth the final vote share. Again, portfolio trading makes sense as the sum of all vote shares is 100. Conditional contracts were proposed by Hanson (2003) to create more complex decision markets. Contracts are paid out either in a binary or index way if a certain preliminary condition is fulfilled (Hanson, 2006). The proposition was made in order to value a nominee under the assumption dependent on who will be the nominees' opponent.
As pointed out by Servan-Schreiber et al. (2004) there are two essential factors driving the prediction market accuracy; knowledge and motivation. One practical way to increase motivation is the promise of winning money. As a consequence the first prediction markets as the Iowa electronic markets (IEM) were set up as real money markets. In order to register participants have to deposit real money. One drawback of real money markets is legal regulations. Real contract exchanges are subject to governmental monitoring and require licenses in most countries. These markets can also be seen as gambling which is prohibited as soon as real money is involved. Gruca et al. (2008) point out that it is an ongoing debate whether real and play money markets behave identical. According

to Servan-Schreiber et al. (2004) and Rosenbloom and Notz (2006) play money markets are as accurate as real money markets. They argue that real money markets may better motivate information discovery while play money markets may yield more efficient information aggregation. Luckner et al. (2008) find that play money markets for the FIFA world cup are about as accurate as betting markets, which are strongly incentivized. Gruca et al. (2008) argue that there is no difference in forecast accuracy as long as there is a lot of publicly available information; otherwise real-money markets perform better. In order to set some incentives in play money markets the usual procedure is to shuffle prizes among participants. There are various winning schemes possible. Luckner and Weinhardt (2007) find rank-order winning schemes lead to the best results in terms of prediction accuracy due to the risk aversion of traders in competitive environments. In contrast to that, Wolfers and Zitzewitz (2006a) state that rank-order tournaments potentially provide an incentive to add variance to one's true beliefs. As fixed payments do not stop traders to be irrationally active in their experiment Luckner et al. (2008) conclude that traders are not only driven by monetary incentives. Spann and Skiera (2003) point out that the motivation to participate decreases if payout dates are too far in future. If the disbursement is limited, the question arises how to motivate traders intrinsically. Christiansen (2007) describes an accurate prediction market forecasting rowing events in the UK with no monetary or prize incentives at all. Furthermore, he speculates that reputation within the rowing community and passion about the game itself generates enough motivation. Cowgill et al. (2009) find when adding *"fun markets"*[2] to serious business related markets that volume in both markets are positively correlated, suggesting that the former might increase participation.

The last design factor, technology, comprises the trading system from an internal market technology perspective and the trading interface. The technology advancements in computer technology and networks have profoundly influenced financial markets and enabled new applications like algorithmic trading. This forced nearly all exchange operators to improve and upgrade their trading sys-

[2]The term "fun" markets refers to market questions whose sole purpose is to entertain participants. They are therefore not directly business related.

tems. Hence, this development has been studied repeatedly. For example the Wagener et al. (2010) focus on the effect of an infrastructure upgrade (which reduced system latency) in the Xetra stock market. They study price efficiency, measured as the pricing gaps between the observed futures prices and their theoretical values based on the underlying cash market. Their results suggest that the system upgrade reduced the pricing gap and improved price efficiency. In the domain of experimental markets, Kolitz (2008) studies the effect of the two different systems using the same auction mechanism on welfare. He finds that the faster system (meet2trade market platform) which allows to process a higher number of bids leads to higher overall welfare.

The second technology aspect is the market interface. As pointed out by Seuken et al. (2011) there is very limited empirical work on decision processes in trading environments with focus on the trading interface. Kauffman and Diamond (1990) highlight the importance of research on behavioral decision making and information presentation effects. They examine how behavioral effects may become operative in screen-based securities and foreign exchange trading activities, where users can choose among information presentation formats which support trader decision making. They present a model to identify where and how information, heuristics and biases might effect decision making in the trading environment.

2.2.2 Market Quality Measures

In the following subsection we detail the tools to systematically analyze prediction markets. To our knowledge there is no stringent methodology to analyze prediction markets. Prediction markets are often quantified solely through the correlation between predictions (derived from prices) and event outcomes. However judging prediction market quality on the basis of forecast precision hinders the understanding of the key success factors. As it is rather untypical to report market measures in the prediction market community; we rely on financial market literature. Due to a missing precise definition of market quality, we provide a list of market quality proxies through which markets can be compared on a more granular level. Following our framework (Figure 2.2) we start by describing activity and liquidity measures. We continue by describing three information measures and close with a short review of the most prominent forecast error measures.

Activity

Activity can be measured using daily measures like the number of trades (trade count), the trading volume (turnover), or the average trade size (see Hendershott et al. (2010) and Zhang et al. (2011)). Trading intensity measures are usually calculated on a daily basis per stock. The first three measures are closely related. An increase in the number of trades does not unconditionally imply an increase in trading volume, since trade sizes also have to be taken into account. While all three measures can be classified as trading intensity, quote updates can be considered as a measure for the mere market activity of traders. The market size is defined by the number of active traders during market activity period. This may range from 15 (in corporate settings) to more than 50 traders which is common in public markets (Plott and Chen, 2002; Christiansen, 2007).

Liquidity Measures

Liquidity represents the transaction cost market participants face to trade. A measure for the liquidity is an asset's ability to be sold rapidly, with minimal loss of value, at any time within market hours (Harris, 2002). We calculate half-spreads rather than round-trip (full) spreads. Quoted spreads are the simplest and most common measure of trading costs and can easily be calculated using trade and orderbook data. All calculations presented below are spreads relative to stock price and are reported in basis points (bps). Let $Ask_{i,t}$ be the ask price for a stock i at time t and $Bid_{i,t}$ the respective bid price. $Mid_{i,t}$ denotes the mid quote then the quoted spread is calculated as follows:

$$Quoted\ Spread_{i,t} = \frac{(Ask_{i,t} - Bid_{i,t})}{2 * Mid_{i,t}} \tag{2.1}$$

Additionally, we separate in quoted spread and quoted spread at trade. The first measure includes all orderbook changes whereas the second is limited to quotes just before a trade is executed.
The effective spread is the spread paid when an incoming market orders trades against a limit order. Let $Price_{i,t}$ be the execution price then the effective spread is defined as:

$$Effective\,Spread_{i,t} = D_{i,t} * \frac{(Price_{i,t} - Mid_{i,t})}{Mid_{i,t}} \tag{2.2}$$

$D_{i,t}$ denotes the trade direction, -1 for a sell and $+1$ for a buy order. The realized spread measures liquidity supplier revenues (Bessembinder and Kaufman, 1997). The Glosten-Milgrom model also highlights that spreads widen if the risk of trading against asymmetrically informed traders is high in order to compensate for losses to the informed traders (Glosten and Milgrom, 1985). The realized spread is calculated with the mid-quote (x) minutes after the trade as follows:

$$Realized\,Spread_{i,t} = D_{i,t} * \frac{(Price_{i,t} - Mid_{i,t+x})}{Mid_{i,t}} \tag{2.3}$$

Estimating the Spread

As in many datasets (e.g. Intrade) the orderbook data is not available for an ex-post analysis, one might estimate the spread. Finance literature suggest several spread estimators. One of the most recent estimate methods was proposed by Corwin and Schultz (2011) (CSS). They derive an estimator for the bid-ask spread based on readily available data of daily high and low prices. The estimator is based on two ideas. Daily high prices (H_t) are almost always buy orders and daily low prices (L_t) are most likely sell orders. The price ratio of high-to-low price is due to volatility which increases proportionately with the length of the trading intervals. Hence they propose to derive an estimate of a stock's bid-ask spread as a function of the high-to-low price ratio for a single two-day period and the high-to-low ratio for two consecutive days ($H_{t,t+1}, L_{t,t+1}$). They define the spread estimator as

$$CSS = \frac{2(e^{\alpha} - 1)}{1 + e^{\alpha}} \tag{2.4}$$

with

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}} \tag{2.5}$$

where β and γ are directly related to the high-low ratios.

$$\beta = \left(ln\frac{H_t}{L_t}\right)^2 + \left(ln\frac{H_{t+1}}{L_{t+1}}\right)^2, \gamma = \left(ln\frac{H_{t,t+1}}{L_{t,t+1}}\right)^2 \tag{2.6}$$

The estimator is easy to calculate and can be applied in prediction market contexts in which only transaction prices are available.
Another proxy for spreads is Amihud's illiquidity (ILLIQ) measure (Amihud, 2002). The measure uses the daily ratio of absolute stock returns in relation to the trading volume in any given period for a specific instrument (i). Where $|R_{it}|$ is the daily stock's absolute return and Vol_{it} is the trading volume during that day, and D_i is the number of trading days during the time period.

$$ILLIQ_i = \frac{1}{D_i}\sum_{t=1}^{D_i}\frac{|R_{it}|}{Vol_{it}} \tag{2.7}$$

Both input variables volume and returns are easily calculated from published prediction market data.

Information Measures

In most prediction markets we can observe the outcome, i.e. the fundamental value of each stock. Therefore, we can ex-post measure the information content of each order. If an order moved the price in the right direction with respect to the final outcome of the stock, it is informed; whereas an order moving the price in opposite direction to the final outcome, it is uninformed. Thus, we present the following score to capture this process:

$$Score_{o_i} = \begin{cases} 1\,if \quad price_{o,i} \leq fv_i \,\&\, o_{type} = BUY \\ 1\,if \quad price_{o,i} \geq fv_i \,\&\, o_{type} = SELL \\ 0\,if \quad price_{o,i} > fv_i \,\&\, o_{type} = BUY \\ 0\,if \quad price_{o,i} < fv_i \,\&\, o_{type} = SELL \end{cases} \tag{2.8}$$

The price of an order o for the stock i is represented as $price_{o,i}$. The fundamental final outcome value of a stock is represented by fv_i. In other words the $Score_{o_i}$ rates an order as profitable or not.

In finance literature the price impact is used as an approximate measure of the adverse selection component of the effective spread. The price impact is the effective spread minus the realized spread and measures the information content of a trade. It approximates the permanent impact of a trade under the assumption that information impacts are permanent and realized at the x-minute mark. Following a trade, liquidity suppliers adjust their beliefs about the fundamental value of an asset depending on the information content of a trade (Glosten and Milgrom, 1985). The simple price impact of a trade is calculated as follows:

$$Price\ Impact_{i,t} = D_{i,t} * \frac{(Mid_{i,t+x} - Mid_{i,t})}{Mid_{i,t}} \tag{2.9}$$

The price impact provides an indication of the information content of a trade.

Again in cases where the full dataset is unavailable, finance literature suggests a promising approximation to quantify the information flow. The probability of informed trading (PIN) represents the implicit risk that a market participant faces when trading with a better informed participant on the direction of the underlying event. Following Easley et al. (2002) we calculate PIN as follows:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b} \tag{2.10}$$

In the Easley et al. (2002) model $\alpha\mu + \varepsilon_s + \varepsilon_b$ gives the arrival rate for all orders and $\alpha\mu$ is the arrival rate for information based orders. With the fraction of informed traders denoted by μ, and the likelihood of information events described by α. The model interprets the normal level of buys and sells per day in a stock as uninformed trade and it uses this data to identify $(\varepsilon_s, \varepsilon_b)$. The days with abnormal levels of buys and sells are interpreted as information-based trading and used to identify α and δ. In order to estimate the model, one only needs the number of buyer- and seller-initiated trades.

One expects the buy and sell trades to be equally likely to be informed ($\varepsilon_s = \varepsilon_b =$

ε) and news are equally likely to be good or bad $\delta = 0.5$. Let α denote the probability that an information event occurs. Informed agents only trade when an information event has occurred. If the information is bad news (δ) uninformed traders buy, if it is good news ($1 - \delta$) they sell. The arrival rate of orders is represented by μ which is assumed to be identical for informed buy and sell orders. With the maximum likelihood method one can estimate the probability that an information event occurs on a given day (α), the probability that an information event is negative (δ) and the order arrival rates of informed and uninformed traders (μ and ε).

Forecast Performance

In binary markets transaction prices typically provide useful (albeit sometimes biased) estimates of average beliefs about the probability of an event (Manski, 2006; Wolfers and Zitzewitz, 2006a). The simplest form of forecast performance evaluation is to calculate the difference between the mean of beliefs and the outcome. However, it is not a one-dimensional concept to judge the forecast quality of probability forecasts. In general, there are three criteria to take into consideration: accuracy, reliability (calibration), and discrimination (resolution). *Accuracy* is a measure of the average distance between forecast and outcome. According to Lichtenstein et al. (1982), *Reliability* or *Calibration* refers to the degree of correspondence between forecast probabilities and actual (observed) relative event frequencies. Whereas *Discrimination* or *Resolution* taps a forecasters' ability to perform better than a simple predict-the-base-rate strategy. Forecasts get perfect discrimination scores when they assign the probability of one to outcomes that happen and probability of zero to outcomes that do not (Tetlock, 2005). *Discrimination* and *Calibration* are best explained in a short example. Let's assume the forecasting goal is to predict the chance of rain at a given time. Let's further assume that on one out of ten days it is raining. A forecast has a high resolution if the prediction points at the day it's going to rain. *Calibration* refers to the overall probability of rain (e.g. 10 %). Thus, a forecaster is well calibrated by assigning a 10% probability of rain to each individual day.

When comparing outcome probability with different underlying variability (e.g.

different domains) one has to take the different uncertainties into account. Nagar and Malone (2011) propose to use the Sharpe ratio to compare reward relative to risk performance. We adopt the Sharpe (1994, 1966)-ratio the following way: in a binary outcome market, let the market return be defined as $R_M = 1 - MAE_M$. With MAE defined as the mean absolute error (see Table 2.1). Note that, in line with the Sharpe-ratio, the higher the positive return the better a forecast. As a benchmark return we use a naive prediction, who bets always 50%. The corresponding benchmark return is therefore $R_B = 0.5$. Furthermore the abnormal return is then defined as the difference between benchmark and market return. Finally, we relate the average abnormal return to the standard deviation of the abnormal returns ($s = \frac{\bar{d}}{\sigma_d}$).

In binary outcome markets it seems intuitive to present accuracy as a hit-rate. If an an outcome has an assigned probability of 50% or more and the outcome comes true, it is counted as a hit otherwise as a false. The hit-rate is than calculated as the percentage of hits. In prediction settings where the economic value of prediction is unclear, and hence it is impossible to derive a error scoring rule it is common to apply the Receiver-Operating-Characteristic (ROC) (Swets, 1988; Zweig and Campbell, 1993). ROC balances the hit rate versus false alarm rate (1-hit rate). This is done by plotting the hit rate versus the false alarm rate over a range of different thresholds that are used to convert probabilistic forecasts of binary events into deterministic binary forecasts. The area under the curve indicated the quality of the predictions, with a perfect predictor scoring of one.

In linear outcome markets, prices do not reflect the probability of an outcome but the market participant's aggregated belief about the fundamental value of the underlying event. Thus, the interpretation of the price is directly linked to the outcome value. There are various ways to generate forecasts from market prices. For example participants can either infer that the $mid_{i,t}$ or the last trading price are the forecast for stock i at time t. In the following sections a market $forecast_t$ refers to the average transaction price on day t.

A first indication about the market outcome is given by the deviation between market prices and fundamental values. In the following the difference between the fundamental value of the stock i (fv_i) and the market $forecast_{t,i}$ represents

Measure	**Definition**
Mean Absolute Error (MAE_t)	$\frac{\sum_{i=1}^{n} error_{i,t}}{n}$
Mean Absolute Percentage Error ($MAPE_t$)	$\frac{\sum_{i=1}^{n} \frac{error_{i,t}}{fv_i}}{n}$
Percent Mean Absolute Deviation ($PMAD_t$)	$\frac{\sum_{i=1}^{n} error_{i,t}}{fv_i}$
Mean Squared Error (MSE_t)	$\frac{\sum_{i=1}^{n} error_{i,t}^2}{n}$
Root Mean Squared Error ($RMSE_t$)	$\frac{\sum_{i=1}^{n} \sqrt{error_{i,t}^2}}{n}$

Table 2.1: Error Measures

the $error_{t,i}$. Studying the error terms we can analyze the forecast properties. to begin with a forecast is unbiased if it has zero mean. Secondly, the forecast is efficient if the error is uncorrelated to the variables (information) used in constructing the forecast. This means the forecast error is unpredictable. Finally, we have to compare forecast errors of different forecast approaches. The comparison of forecast approaches depend on a user's cost function (sometimes also referred to as scoring rules) (Clements and Hendry, 2002). In principle, costs can be attached to a forecast bias and error variance. These costs will depend on the purposes of the forecast. Thus, cost functions define how forecast properties should be weighted when comparing different forecasts. For example, if (squared) bias and variance are combined one-for-one, we obtain the forecast's mean square error (MSE). Table 2.1 presents some commonly-reported forecast error measures. A detailed discussion of the presented measures can be found in (Fildes and Stekler, 2002).

A problem with the error functions defined in Table 2.1 is their comparability. If applied to series with different units or different baselines, comparing the error terms leads to misguided results. Hence, one needs to measure the relative accuracy. One idea is to normalize the forecast error by a naive benchmark forecast. This measure is called Theil's U statistic (Leitch and Tanner, 1991). It is generally defined as

$$Theil's\, U = \frac{RMSE_{forecast}}{RMSE_{naive-forecast}}. \tag{2.11}$$

The forecast in question is thereby compared with a naive forecast derived by mere chance (e.g. random walk or a simple AR(1)-Model).
If the value of U is less than 1 the forecast is *relatively* good. If U equals one, the forecast is just as good as the naive forecast. U-values greater than one indicate that the forecast is of little use. Theil's U is specifically suitable for the measurement of the forecast quality since it does not only make the forecast quality assessable between the respective forecasts, but represents an ultimate measure for forecast capability as well.
In order to test for weak form forecast efficiency in which forecasts contain all available information at the time the forecast is generated, we adapt a test by Nordhaus (1985). Weak-form forecast efficiency means that all forecast revisions and errors should be uncorrelated with past forecast revisions. Expressed differently revisions and errors should follow a random walk. In the following a revision is defined as:

$$revision_{i,t} = forecast_{i,t-1} - forecast_{i,t} \tag{2.12}$$

To test for correlation we use the following OLS regression:

$$rev_{i,t} = \alpha * rev_{i,t-1} + \beta * rev_{i,t-2} + \gamma * rev_{i,t-3} \tag{2.13}$$

Note that forecast efficiency differs from market efficiency by Fama (1970, 1991) in the sense that it does not test for correlation on a trade by trade basis but on an aggregated daily level.
In order to quantify the information contained in comparable forecasts we propose to use the Fair-Shiller Model (Fair and Shiller, 1989). The information contained in one forecast compared to another forecast, from a different source, can be assessed using a regression of actual values on predicted values generated from the forecasts. Let $forecast_{t,com}$ denote a comparable, in terms of unit and

time horizon, forecast to the market $forecast_{t,market}$, then the estimates in the following regression represent the information contained in each forecast.

$$actual\ outcome = \alpha * forecast_{t,com} + \beta * forecast_{t,market} \qquad (2.14)$$

If both forecasts are just noise, a, b are zero. If both forecasts contain information but the information in one forecast is completely contained in the other forecast, hence one forecast does not contribute additional information, then one but not both estimates should be nonzero. Fair and Shiller (1990) show that the regression is useful in comparing different forecast models.

Combining Forecasts

Combining forecasts can reduce error in several ways. A combined forecast is likely to be more accurate than a typical forecast of an individual component, because biases associated with the data and methods used in various forecasts are likely to differ. Bates and Granger (1969) find that combining two forecasts reduces the error by 11% on average. In line with this early finding in a meta-analysis of 30 studies, Armstrong (2008) finds that combined forecasts have a 12% lower error than the average forecast component. However, literature also points out why combined forecasts are often not applied. For example, Larrick and Soll (2006) show that people often hold incorrect beliefs about averaging, falsely concluding that the average of two forecasts would be no more accurate than the average of both forecasts. Previous findings suggest that forecasts should be weighted equally, unless there is strong prior evidence that supports differential weights. Hogarth (2006) points out that it is difficult to accept the idea that simple models can predict complex phenomena better than complex ones.

2.2.3 Markets for Economic Outcomes

Markets for macroeconomic variables have been used since the 80s. The Coffee, Sugar and Cocoa Exchange established a futures market for the Consumer

Price Index allowing traders to hedge on inflation. The market was closed due to low interest (Mbemap, 2004). In 1993 Robert Shiller argued for the creation of 'Macro Markets' which would allow a more effective risk allocation (Shiller, 1993). In 2002 Goldman Sachs and Deutsche Bank set up the so called "Economic Derivatives" market tied to macroeconomic outcomes such as ISM Manufacturing, change in Non-Farm Payrolls, Initial Jobless Claims and Consumer Price Index (Gadanecz et al., 2007). The traded contracts are securities whose payoffs are based on macroeconomic data releases. The instruments are traded as a series (between 10-20) of binary options. For example a single data release of the retail sales in April 2005 was traded as 18 stocks. In order to maximize liquidity the market operators used a series of occasional dutch auctions just before the data releases instead of the more common continuous trading on most financial markets. Thus the market provided hedging opportunities against event risks and a short horizon market forecast of certain economic variables. By analyzing the forecast efficiency Gurkaynak and Wolfers (2006) find that market generated forecasts are very similar but more accurate than survey based forecasts. One must note, that the Bloomberg survey forecasts are published on Fridays before the data release, whereas the auction was run, and the forecast was generated, on the data release day.
In an attempt to forecast inflation changes in Germany, Berlemann and Nelson (2005) set up a series of markets. The markets feature continuous trading of binary contracts. In a second field experiment Berlemann et al. (2005) used a similar system in order to aggregate information about inflation expectations in Bulgaria. All in all, the reported forecast results in both experiments are mixed and not too promising. Besides the low number of forecast events both experiments suffer from low public interest resulting in illiquid markets.

A Case for a New Market Design

As detailed in the last section, previous research focuses on binary contracts. However, the standard approach reaches its limits if the number of outcomes is very high or even infinite. For instance, in a market to assess GDP growth, possible outcome ranges from -100% to infinity. A common work-around is to set arbitrary intervals over the range of possible outcomes and trade each interval as an individual stock. The market operator faces two decisions in such a setting. First, she has to pre-estimate a reasonable range of possible outcomes. Secondly, she has to set corresponding intervals. E.g. if the pre-estimated window for GDP growth is between 0% and 5% then the market operator still needs to define the number of intervals. A fixed interval size already limits the accuracy of the prediction and the choice of range might bias a prediction market's results. In the GDP case mentioned, market participants have the choice of six answers (six different stocks) in 1% intervals. Even if market participants predict the right interval, such a prediction market would still yield inaccurate forecasts. Additionally as it is desirable to forecast not only the next upcoming period but longer horizons, the number of needed contracts rises. Using binary contracts with 1 %-intervals, five indicators and three periods per indicator would lead to a minimum of 60 contracts. The high number of contracts would result in low liquidity and eventually diminish the forecast accuracy (Brenner et al., 1999; Abramowicz, 2004).

Analyzing the "Economic Derivatives" market dataset, Sonnemann et al. (2008) find a bias which they called *"partition-dependence"*. They show that by arbitrarily setting intervals on the state space the market operator influences the judged likelihood. Thus, all previous markets suffer from a bias induced by the market operator.

It is well known that people have difficulties understanding and using probabilities (Tversky and Kahneman, 1974). In order to reduce the complex task of accessing probabilities people rely on a set of heuristics which sometimes lead to serve and systematic errors. In particular, people underweight outcomes that are merely probable in comparison to outcomes that are obtained with certainty (Camerer and Lowenstein, 2003). In (betting) markets this leads to the long-shot bias. Near certainties are undervalued whereas low probabilities are overvalued (Berg and Rietz, 2006; Snowberg and Wolfers, 2005). Thus, Wolfers and Zitzewitz (2006b) advised caution interpreting the prices of low probability events.

It seems clear that the representation of more or less continuous outcomes through intervals does not produce ideal results. The benefit of representing an event with a range of all-encompassing stocks is out-leveraged by the hassle of trading a large number of stocks. The market design, proposed in this thesis, tries to circumvent the presented problems by representing events as linearly paid out contracts.

2.3 Summary

In the previous section presented our research foundations. By describing the market engineering approach, we identify three open questions. To begin the market user interface as the first point of contact with market participants has yet not been considered as influencing participant behavior. This gap will be addressed in chapter seven. Secondly, the idea of rapidly developing and improving markets (*Continuous Market Engineering*), seems promising in the fast changing e-commerce world. However, there has been no work done on applying and empirically evaluating the method.
Closely related is the topic of market evaluation; until now there is only a limited amount of literature on measuring market quality. In the second half of the previous chapter we addressed this issue. By summarizing previous work we detailed a prediction market framework with a strong focus on how to evaluate prediction markets. The framework combines and applies both finance and forecasting literature. The market performance measures will be applied in chapter four.
Looking at previous markets for economic indicators, we found promising but mixed results regarding market accuracy and liquidity. Moreover the previously employed digital contract structure seems to have drawbacks when used in continuous outcome settings. This leads to the question if there are different ways to design efficient markets for economic variables.

Chapter 3

EIX - The Economic Indicator Exchange

ACCURATE and reliable forecasts of future short- and long-term economic developments are a crucial competitive factor for companies, regions and countries and an important foundation for political decision making. Hence, it does not come as a surprise that the prediction of future business cycle developments is one of the most extensively pondered subjects in economic research. Over the past decades a broad variety of technical, statistical as well as qualitative methods have been developed to foresee economic trends.
It is well a known that traditional economic forecast models lack the necessary accuracy (Clements and Hendry, 2002).
As Oller and Barot (2000) point out the quality of forecasts has generally not improved over the past 40 years despite massive progress in statistical methodology and computer technology. In fact quantitative, technical methods have proven to fail regularly when major changes to the general economic environment and paradigm shifts appear (Osterloh, 2008; McNees, 1992; Schuh, 2001). Yet another issue is the reliance of the current forecasts on expert input. Experts are prone to biases and political influence and generally do not perform better than novices in forecasting future events (Armstrong, 2008). Due to the reliance on personal judgments, forecasts have been found to exhibit a bias towards optimism (Batchelor, 2007). In Germany forecasts are produced by numerous institutions and released on periodical basis. Moreover, forecasts vary in time horizon

and definition. Thus, economic agents (e.g. decision maker) relying on these forecasts might find it difficult to aggregate the various forecasts and come to a confident appraisal. To conclude, up till recently there is no central mechanism to combine individual forecast to from a common informative basis.

3.1 Introduction

In October 2009 a play money prediction market, specifically designed to forecast economic indicators such as GDP, inflation, investments, export and unemployment figures in Germany, was launched. The market called Economic Indicator Exchange (EIX)[1] was created in cooperation with the leading German economic newspaper "Handelsblatt". The cooperation aims at reaching a wide and well informed audience interested in financial markets and economic development. We thus expect no problems understanding the indicators and the concept of trading. The market is publicly available over the Internet and readers are invited to join. The registration is free and requires besides a valid email address just minimal personal information.
When starting the project we intended to maintain it for one year. However due to its success we extended it for another two years. After the first year (named version one, or round one) we slightly adapted the system for the second year (version two, or round two). The third year is not covered here.

3.2 Market Design

The market design features a continuous double auction without designated market maker. Participants are allowed to submit marketable limit orders with 0.01 increments through a web-based interface. After registration participants are endowed with 1,000 stocks of each contract and 100,000 play money units. We propose to represent continuous outcomes with one stock and define a linear payout function. Contracts for each economic indicator are paid out according

[1] www.eix-market.de or eix.handelsblatt.com

to equation 3.1.

$$p = 100 + \alpha \times (\frac{I_{t0} - I_{t-1}}{I_{t-1}})\ with\ \alpha = 10 \qquad (3.1)$$

A contract is worth: 100 +/- α times the percentage change for an indicator in play money (e.g. a change of 2.1 % results in a price of 121). We set α to 10. Therefore the representable outcome events range from -10 % to infinity. To represent the whole outcome range from -100%, α could be set to one. Previous work indicates that market participants find it difficult to estimate minor changes in the underlying (Stathel et al., 2009). Hence, we propose to scale the minor changes to a certain level. Looking at historical data there were no events where the German GDP dropped 10 % per quarter. The rationale for setting α to 10 was the deliberation that participants find it more intuitive to enter integers in order to to express reasonable accuracy. Additionally, German statistical data releases rarely come with more than one decimal.

Table 3.1 summarizes the economic variables tradable on the market. Due to the payout function and the selection of the corresponding units; all stock prices are expected to roughly range between 50 and 150. Therefore participants could similarly gain by investing in specific indicators. To facilitate longer forecast horizons every indicator is represented by three independent stocks each representing the next three data releases (t_1, t_2, t_3). As a consequence the initial forecast periods vary between one month for monthly released indicators up to three quarters for quarterly released variables. The day before the release date the trading in the concerned stock is stopped. Finally, stocks are liquidated according to the payout function defined in equation 3.1. As soon as the trading in one stock stops a new stock of the same indicator (e.g. t_4) is introduced into the market. This means that participants received 1,000 new stocks of the respective indicator. All in all, participants are able to continuously trade 18 stocks at all times.

3.2.1 Contracts

The indicators are a mix of leading (forecasting the economy, e.g. Investments) and lagging (describing the state of the economy, e.g. Unemployment numbers)

Indicator	Unit	Data release cycle	N	Pay-off function
Exports (V1)	$rel. - Changes_{t-1}$	monthly	12 (-)	$100 + \alpha \times (\frac{I_{t0} - I_{t-1}}{I_{t-1}})$
Exports (V2)	Billion(abs.)	monthly	- (12)	$ABS(Number) + 30$
GDP	$rel. - Changes_{t-1}$	quarterly	4 (4)	$100 + \alpha \times (\frac{I_{t0} - I_{t-1}}{I_{t-1}})$
Ifo Index (V1)	$abs. - Changes_{t-1}$	monthly	3 (-)	$100 + \alpha \times (I_{t0} - I_{t-1})$
Ifo Index (V2)	Points (abs.)	monthly	- (12)	$ABS(Points)$
Inflation	$rel. - Changes_{t-12}$	monthly	11 (12)	$100 + \alpha \times (\frac{I_{t0} - I_{t-12}}{I_{t-12}})$
Investments	$rel. - Changes_{t-1}$	quarterly	5 (4)	$100 + \alpha \times (\frac{I_{t0} - I_{t-1}}{I_{t-1}})$
Unemployment	Million (abs.)	monthly	12 (12)	$100 + \frac{ABS(Number)}{100.000}$

Table 3.1: Economic variables and their payoff functions. The Ifo Index was introduced in August 2010. For the second round the Export and Ifo pay-off function were changed. N denotes the number of payouts in the first and (second) round.

economic indicators.

One should note that contracts are liquidated based on the first data release (preliminary estimate). However, one mayor problem with macroeconomic data is its poor quality and its slow availability (Croushore and Stark, 2001). For instance, export data is typically available only after a quarter, and the subsequent revisions (over the next 2 years or so) can be sizable. There are much less, if any, revisions in data on unemployment and inflation data. Additionally to the revision problem, one has to be very careful about the contract specification and the data adjustments used. We base our payouts on the seasonally-adjusted, calendar-adjusted data release of the official statistical agency[2]. For Exports, Investments and GDP the data is also price-adjusted. The exact definitions are given on the web page. The selection and precise definition follows what the media usually reports in a more casual way[3].

[2]www.destatis.de, the agency uses the ARIMA.X12 to adjust the data.

[3]Before the experiment started we had several discussions with media experts from Handelsblatt and economists from the Institut der Deutschen Wirtschaft, Köln to ensure the correctness of the definitions.

3.2.2 Short Selling

Since we do not have complementary stocks which would allow participants to artificially create stocks by buying a bundle, a challenge of our design is how to inject stocks in the market in the first place. We employ two potential approaches. To begin with, we endow all traders with an initial stock portfolio. Secondly, we allow traders to sell stocks without actually owning a stock, known as short selling in financial markets.

Comparable to financial markets, short selling needs some restrictions in prediction markets. If traders want to buy stocks they are limited through their initial cash endowment and subsequent cash changes. Since short selling involves selling stocks that traders do not possess there is no natural, corresponding limit for selling stocks. In short selling enabled prediction markets there is the need for a short selling restriction equivalent to the budget constraint for buying stocks.

There are several possibilities to implement short selling in prediction markets and to implement short selling constraints. The market operator could generally function as a credible lender of stocks. It is important to make sure that market participants are liquid enough to meet their short obligations once trading in a specific stock ends. One basic approach is to implement a fixed constraint through the number of shares that a trader can go short. Other possibilities include dynamic implementations which are based on potential losses a trader would have to realize in extreme cases through short selling. In our opinion there is no general rule how to implement short selling constraints. However an approximation of the maximum number of borrowed stocks is given by equation 3.2.

$$Max.\ Number\ of\ Borrowed\ Stocks = \frac{Portfolio\ Value}{Max.\ Loss} \tag{3.2}$$

In the case of the EIX market the maximum loss is practically limited through the EIX market design. The highest expected stock price is 200 (e.g. a 10 % growth rate per quarter). Thus the maximum expected loss is selling a stock for 0 and a payout of 200. Following these practical considerations we set the $Max.\ Loss$ to 200. Hence, the maximum number of borrowable stocks varies for each market participant. The actual borrowing takes place while trading. Stocks

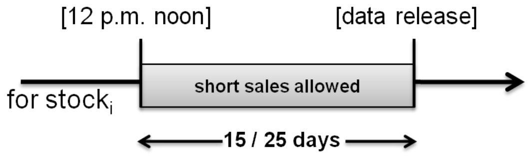

Figure 3.1: Allowing short sales

in which short selling is allowed are marked by a small icon on the web-interface. If a trader sells more stocks than she owns, the market (operator) automatically acts as the lender. Consequently, borrowed stocks are interpreted as liabilities in the portfolio value calculation. The last aspect considered is the timing of the short sale allowance. Short selling was allowed in single stocks from March 15th till the 31st of October 2010. In order to test the effect of short selling on market forecast accuracy we introduced short selling for each stock separately. Short sales were allowed at noon, 15 respectively 25 days before the data release (Figure 3.1). This allows us to analyze the effect for each individual stock.

3.2.3 Incentives

As mentioned, the market is a free to join play money market. In order to motivate participants intrinsically we provide two interface features; traders can follow their performance on a leader board and they can form groups with others to spur competition with friends. Previous research in the field of prediction markets has shown that play-money markets perform as well as real-money markets predicting future events (Wolfers and Zitzewitz, 2004; Servan-Schreiber et al., 2004). Due to the legal restrictions on gambling the EIX prediction market has to rely on play money. To increase participants' motivation and to provide incentives to contribute information we hand out prizes worth 36,000 Euro. In order to be useful, an accurate prediction must be determined well in advance of the actual outcome. From a forecasting perspective it seems meaningless to run a market where one obtains the prediction just before the actual outcome occurs. This sounds obvious, but it is actually quite difficult to achieve, because traders want to know how their investment turned out, fairly quickly. As we try to forecast longer periods the incentive scheme has to address this problem. So the

incentives are divided in two parts (a) monthly prizes and (b) yearly prizes. The 8 yearly prizes (total value 10,000 Euro) were handed out according to the portfolio ranking at the end of the market. The monthly prizes are shuffled among participants who fulfill two requirements for the respected month: (i) they increase their portfolio value and (ii) they actively participate by submitting at least five orders. Both incentives are clearly communicated through the web-interface. For the yearly prizes the leader board indicates the current status of all participants. The monthly winning status is displayed individually just after each login.
Due to a lower number of sponsors, the amount of prize money was reduced in the second round. We handed out three prizes worth 1,030 Euro per month; 12,360 Euro overall.

3.3 The Trading Software

3.3.1 IT Software Architecture

In addition to the key design elements of the EIX prediction market described in the previous section one also has to design the web-based trading software as well as the facilities handling information about the traders' accounts, the order matching and quote updates from a technical point of view. The EIX prediction market software is an advancement of two previously run markets (Stathel et al., 2009). The system is implemented in Grails[4]. It features a modularized architecture in order to keep it easy to maintain and expendable by services and functionality. Due to the previously unknown number of users the software platform has to be scalable. Figure 3.2 summarizes the whole system from three perspectives; IT-infrastructure, application logic and the core order management. The *IT-infrastructure* is provided by the Forschungszentrum Informatik, Karlsruhe (FZI), it consists of three physical servers; a Squid reverse proxy, caching the static pages, a designated PostgreSQL server for the database and a tomcat application server; running the application logic. The *application logic* has been set

[4]www.grails.org

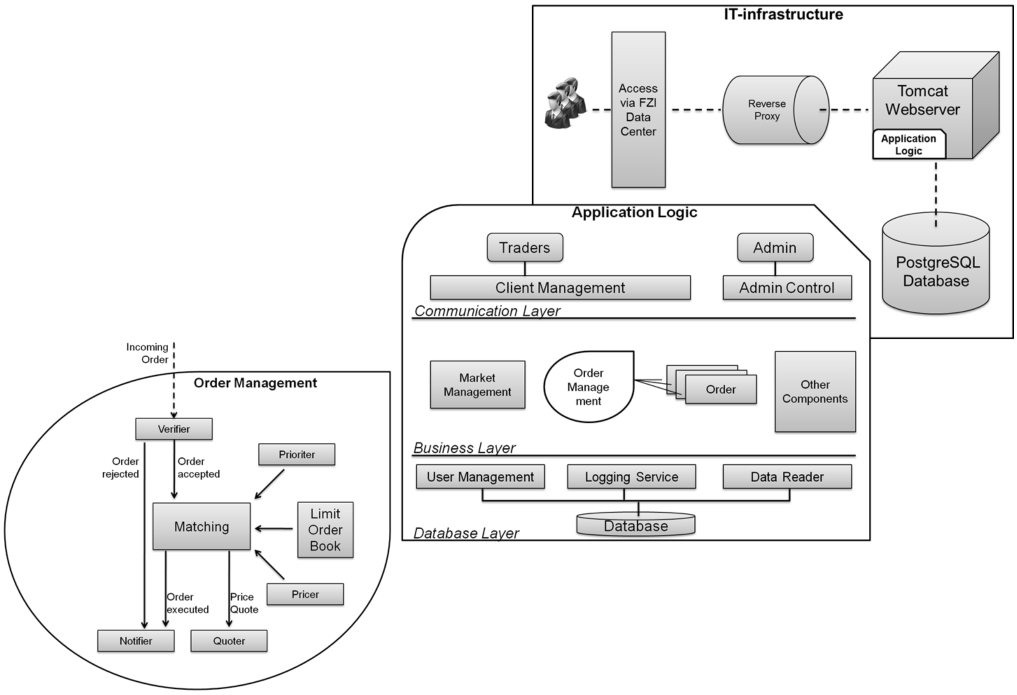

Figure 3.2: Software system

up following the model-view-controller concept. Therefore, it is separated in three layers; one handling the external communication e.g. the website presentation, one for the internal database querying and finally one running the core order processing. As the core element the *order management* processes all incoming orders. The EIX market employs the commonly used trading mechanism; the continuous double auction (CDA). In a CDA, known e.g. from the Deutsche Börse system Xetra, traders submit buy and sell orders which are executed immediately if they are executable against orders on the other side of the order book (Madhavan, 1992). If orders are not immediately executable, orders are queued in an order book and remain there until they are matched with a counter-offer, or are actively deleted by either the market operator or the submitting participant. Orders are executed according to price/time priority, i.e. buy orders with a higher limit and vice versa sell orders with a lower limit take priority. In case several orders were placed with the same limit price, the orders which were submitted earlier are executed first. One of the main advantages of using a CDA is the fact that markets with a CDA pose no financial risk for market operators as they are a zero-sum game. Moreover, the CDA allows for continuous information incorporation into prices and consequently traders are capable of quickly reacting to events.

3.3.2 Trading Interface

The default trading interface is displayed in Figure 3.3. On the first arrival, the participant only sees the trading mask (box upper left side marked with 0). It contains the necessary options and fields to submit an order. First the user decides whether to buy or sell the selected stock. Changing the order type adapts the trading interface. For example, the small icon changes for a buy order to a shopping cart which is going to be filled. The label next to the limit price changes from lowest price to highest price.

In the third row the participant then specifies the limit price. She can change the limit price three ways. Firstly by inserting the number directly (with a maximum of two decimals) or secondly by typing in her prediction, which is then translated into a limit price. Or thirdly by using the arrows to increase or decrease the prediction and limit price by one increment.

According to Thaler et al. (2010) the default settings in systems matter. Hence, the defaults were chosen purposefully not to distract or anchor participants to certain values. The order type is set to sell. Previous experience indicates that participants are slightly biased towards the buy action (Stathel et al., 2009). The order size is set to 100 which represents 10 % of the start portfolio and the default limit price (prediction estimate) is set to 100 (0 %). In the last row the user has to specify the number of stocks being bought or sold. As additional information, the system provides the current portfolio for a sell order and the highest number of stocks a user can buy with the current specified limit price. Moreover, a short description of the market comprising the respective payoff function was shown as part of the trading screen.

Participants are able to customize their trading interface individually. By clicking the small arrows seven information panels open and close. In the default setting, only the trading mask and the seven headlines are visible. Hence, participants have convenient access to the order book with 10 levels of visible depth (1), the price development (2), the account information (3) and market information (4) such as the last trading day. As additional information the Handelsblatt provides access to an up-to-date economic news-stream (5) and finally the indicator's last years performance (6) is displayed. In the second version, we added a panel to display a list of previous forecasts (7). The list contains all orders of the selected product submitted by the currently logged-in participant. Furthermore participants are able to directly cancel previous orders. After each submitted

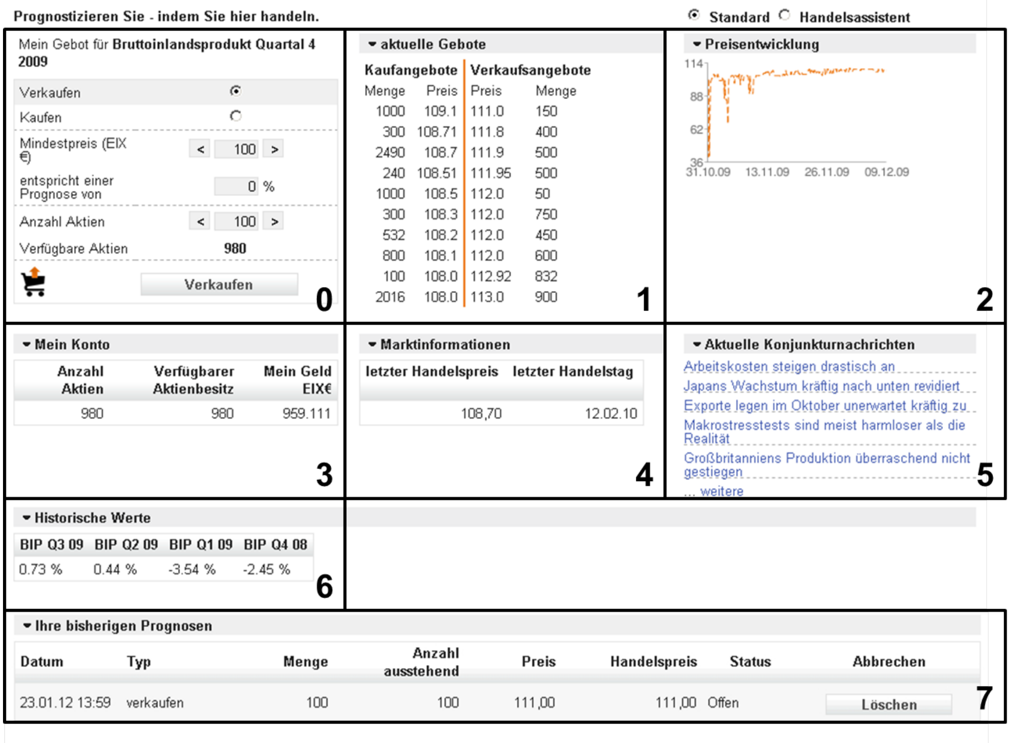

Figure 3.3: Default trading screen

order the chosen interface is saved per user. When the user returns, the system opens the previously used interface elements on default. The advantage is twofold; users have a convenient option to customize their trading experience and, we can assess which self-selected information pieces may have influenced the participants' decision processes. Moreover, we did not have to form groups with different interfaces and assign users to certain groups. This setting would possibly create an unfair experience. Additionally to the default trading inter-

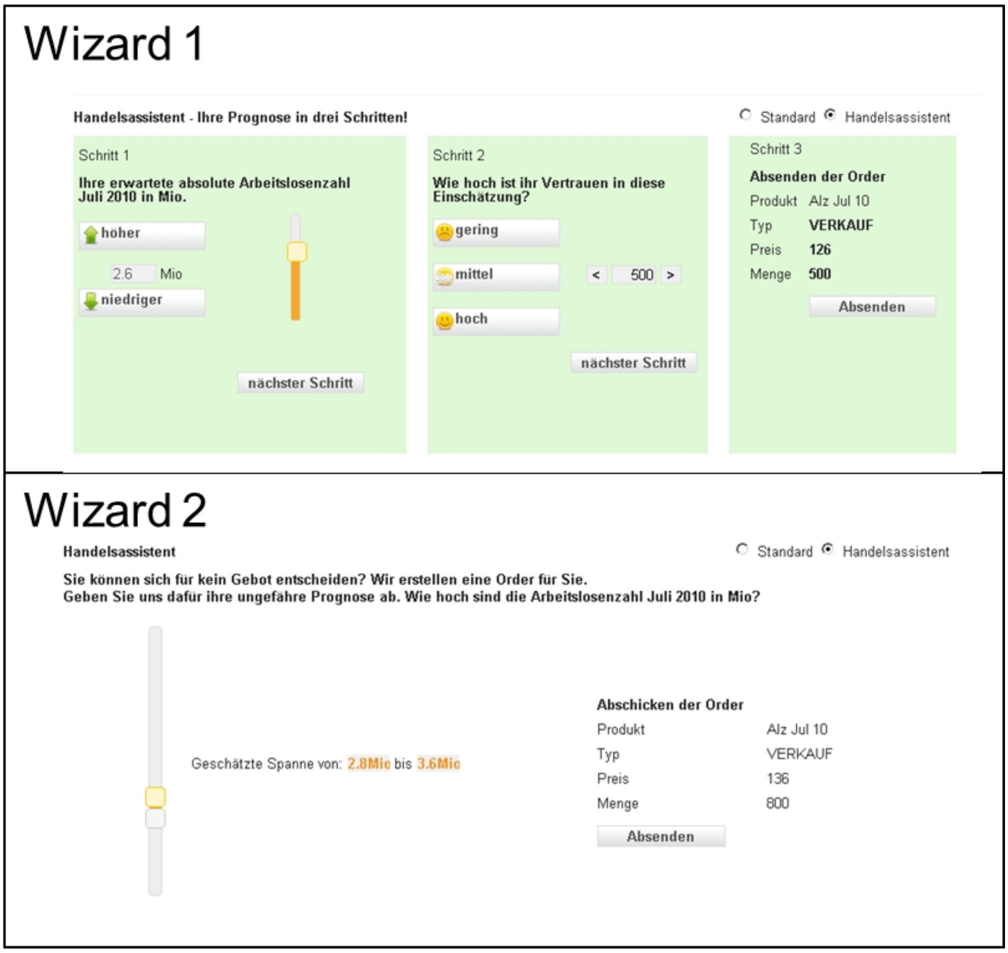

Figure 3.4: Two trading wizards

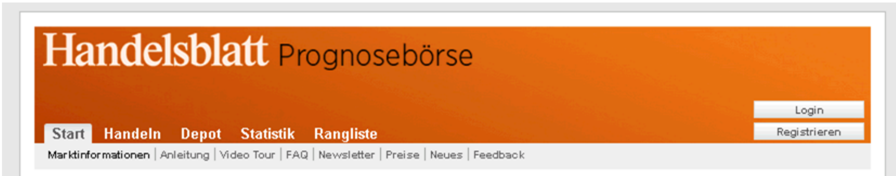

Figure 3.5: Navigation bar

face, participants have the choice to switch to a trading wizard guiding their trading decisions (Figure 3.4). In order to test for the interface influence on trading performance we designed two different wizards marked with $Wizard_1$ and $Wizard_2$. Participants are randomly assigned in one of two groups with access to one of the two different trading wizards. Interface $Wizard_1$ is designed as a three step trading wizard, with three (green) boxes appearing in sequence. In the first step participants indicate if they believe the prediction to be higher or lower than the current market forecast. In the second step they are asked about their confidence in their prediction. The third box just displays the generated order.
Interface $Wizard_2$ simply asks the participant to indicate a prediction interval with two handles. On the right hand side an order is automatically generated depending on the current orderbook and the distance between lowest and highest indicated prediction value. It is noteworthy that both wizards provide far less information than the default interface. In terms of Seuken et al. (2010) interface type $Wizard_1$ can be considered as a weakly hidden market interface, whereas type $Wizard_2$ hides the market completely . Figure 3.5 displays the full EIX menu structure. The portal also provides more information on the prizes traders can win, the operational principle of the prediction market including a video tutorial and frequently asked questions, as well as an up-to-date news stream related to the German economic development. The second menu item holds the available account information for individual traders including the number of shares held in each contract, the balance of the cash account, the total value of their deposit, a list of outstanding buy and sell orders, as well as a list of trades. The last item leads to a ranking of all the traders sorted by their deposit value, i.e. the balance of their cash account plus the value of the contracts they held at the specific point in time.

3.3.3 Feedback Mechanisms

In our market setting we distinguish between four types of feedback:

- Interface feedback
- Market based feedback
- Forecast performance feedback
- Participation feedback

Figure 3.6: Feedback on default trading screen

The first feedback type is directly communicated through the *trading interface*. If participants enter a limit price another field displays the related prediction for that price. Vice versa, participants can change their prediction and see that the related price adapts automatically (See figure 3.6, F). This feature helps to communicate the complex contract design previously described.

Market based feedback is communicated on various levels. First of all stock prices reflect the current aggregated belief of other market participants. Moreover the orderbook displays the current market confidence about a certain event (e.g. with a high spread). Finally, after contracts are liquidated, participants can easily follow their own contribution in relation to their peers. This confronts forecasters with their own forecasting performance. Additionally as good forecasters increase their portfolio value they gain more weight over market run-time.

Aggregating trading success in order to create a portfolio-based ranking is the standard way of giving *forecast performance* feedback. In order to separate between trading performance and forecasting performance the platform offers additional performance-feedback; the so-called "EIX-score". It is calculated based

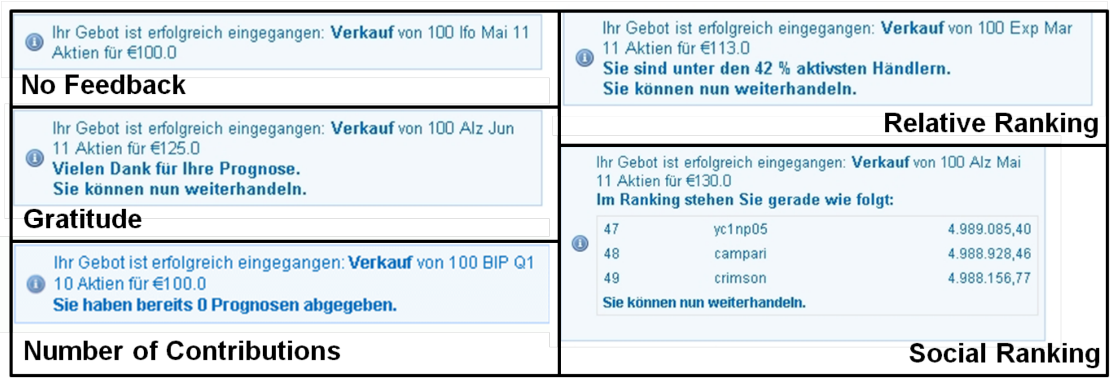

Figure 3.7: Participation feedback

on the observed outcome, i.e. the fundamental value of each stock. If an order moved the price in the right direction with respect to the final outcome of the stock, it is informed; whereas an order moving the price in opposite direction to the final outcome price, it is uninformed. Therefore we can ex-post measure the information content of each order. Combining the information content with the size of an order and aggregating all individual contributions enables us to calculate a forecast based performance ranking. *Participation feedback* is communicated directly after a user submits an order. Mazarakis et al. (2011) show that feedback and recognition increases participants' intrinsic motivation to contribute to public goods projects. Following previous work we created six treatment groups. Figure 3.7 depicts five of them. The 6th. group is a randomized treatment group which gets randomly one of the five feedback types after each submitted bid. All groups get a confirmation that their order was successfully submitted. We call this the no feedback group. The other feedback treatments are:

- *Gratitude* which is a simple "Thank you"-message displayed after a contribution.
- A *Historical Reminder* which shows the user her number of contributions in order to think about her own past contribution behavior.
- A *Relative Ranking* displays the contribution frequency compared to peers.
- A *Social Ranking* feedback, which illustrates a ranking within a group of users who created similar contributions.

N= 1227	1 (very good)	2	3	4	5	6 (unsatisfactory)
Economic Knowledge	10.11 %	40.10 %	34.31 %	9.78 %	3.59 %	2.12 %
Market Knowledge	10.19 %	35.21 %	34.96 %	11.65 %	5.79 %	2.20 %

Table 3.2: Participant statistics

3.4 Market Participants

The registration at the EIX-market is free and requires besides a valid email address just minimal personal information. Upon registration we asked participants to self-assess their market-knowledge and their knowledge of German economy. Furthermore, participants indicate if they are working in areas related to exchanges or economic forecasting; with 37.16 % indicating that they do. Table 3.2 summarizes the first participant statistics. The user input is highly correlated ($\rho = 0.74$), meaning that particpants indicating high knowledge in the market domain do the same in the economic domain. In further analysis we combine the scales to a confidence proxy. As a requirement for participants to qualify for prizes, they had to provide information about their address and age. Using the name and email-address to identify gender, it seems that most participants are predominant male (93 %) and on average 53 years old. Finally, in order to spur competition between friends, the first version of the portal offered the possibility to create groups. Group members could invite acquaintances. Within the group, members were able post messages and saw a group-internal ranking, based upon the overall portfolio ranking. Additionally the group portfolio is summed-up and a separate group-ranking is displayed, with all groups ranked according to their performance. During the first round only 44 participants formed 12 groups.

3.5 Experimental Timeline

Figure 3.8 displays the EIX market development. Even though the experiment was intended to stay relatively stable over the time horizon, some changes were necessary to keep participants interested and active.

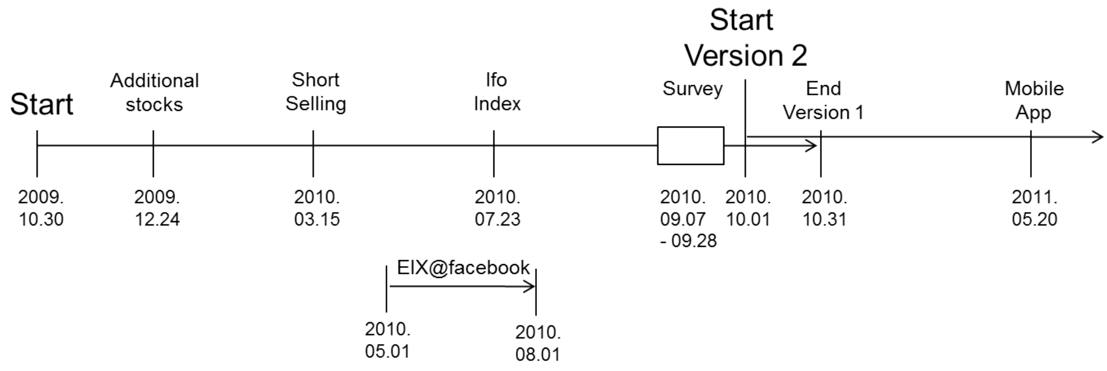

Figure 3.8: Experiment timeline

To begin with, on Christmas eve 2009, we handed out 500 additional units of every tradable stock as a Christmas present. As with all market changes, we announced this action in advance through the weekly newsletter and directly on the web site.

We received feedback from market participants that stocks are overvalued. However, with no more stocks to sell they had no opportunity to correct the price. Hence, short selling was allowed in single stocks beginning in March 15th 2010. Similarly we received feedback asking for more indicators to be traded. In response we introduced the Ifo business climate index in July 2010. This means that participants received three times 1,000 stock units for the following three periods.

The first round (version one) of the EIX prediction market ended on the 31st October 2010. In parallel we started the second round (version two) on October the first. Every market participant who registered for the first version was automatically transfered to the second round. No new registration was required and the website layout, web-address and institutional setting remained the same. However several market changes were realized.

First of all, we changed the payout function of two indicators. The Ifo index was directly related to the price ($Price = Ifo - index\ (points)$). The intention was to make it easier for participants to translate a prediction into a limit-price. As the predictions for the export indicator were performing the worst compared to the other indicators we changed the payout function to improve the

forecast quality. The new payout function was the *Exports* in billion minus 30 ($Price = Exports(billion) - 30$). The long-term average export value is around 70 billion per month. For the function we scaled the average (70) to 100 and added/subtracted the difference to the average.
As the positive impact of short selling on market forecast quality became obvious, we extended the time horizon in which short selling is allowed. Five days after a stock is introduced (handed out to participants) participants are allowed to short-sell the respective stock.
In the second version the amount of prize money was reduced due to a lower number of sponsors. We handed out three prizes worth 1,030 Euro per month. Due to a missing year-end prize we adapted the incentive scheme. Traders are rewarded on a monthly basis and not a yearly basis. To successfully communicate the winning status we generate a score for every matched order.
As described previous Chapter, we can measure the information contained in each order. If an order moved the price in the right direction with respect to the final outcome of the stock, it is informed; whereas an order moving the price in opposite direction to the final outcome price, it is uninformed. We use the score definition in equation 2.8 and weight the score by the order quantity submitted. As every order is linked to one market participant, we can aggregate the score to determine the monthly individual performance.

$$Score_{p,m} = \sum_{o=1}^{n} score_{o,i} \; for \; i \; paid \; out \; in \; m \tag{3.3}$$

The monthly score per user (p) is given by the sum of all users' orders for stocks which are paid out in the respective month (m). The top three market participants according to the score of each month were rewarded with one prize each.
Part of the preparation of the second version was a review of all functionalities. Within this review we realized that only a small fraction of participants used the feature to form groups. It seems that participants did not find it useful to create groups without being explicitly incentivized to do so. In order to streamline the platform, we removed this feature in the second round.
The trading interface was only slightly adapted. On the default trading interface

a new panel displaying previous orders was added (Figure 3.3, panel 7). The trading wizards remained the same. Regarding information logging, we added a timer, measuring the time a user takes to submit an order.
In order to prohibit supposedly unintended user input we added a verifying heuristic. The heuristic is based on the two aspects of long time historical average for an indicator and the previous average forecast of the user. For example, if the submitted limit price deviates too much from the indicator's historical average or if the limit price deviates beyond a threshold from the trader's average limit price the user is sent to another page. This page (Figure A.1) displays either the historical average or the trader's average limit price as a benchmark. The trader can than either abort the order, change the limit price or just continue to submit the order. In order to potentially measure the effect of such verifying rules we record both the initial order as well as the finally submitted order.
As we received positive feedback and participation was still high in end of version two, we continued to run the EIX-platform for another year. For the third version we adapted the trading wizard and the website presentation was redesigned (Figure A.3). However to report a full data-set, the following sections will only cover version one and two.

3.5.1 EIX@facebook

In a parallel experiment we integrated an identical market as an application inside the social network facebook. For three month from May 2010 till August 2010, we invited students to trade in this market. As incentives (500 Euros) and advertisement were kept low-key, only 50 traders signed up. This allows us to compare the forecasts generated by thin and thick markets. Additionally facebook provides insights into the trader connections which can be linked to individual trading behavior (Teschner and Rehm, 2011). Due the small size of the parallel market and it's short duration there were no effects on the main EIX-market.

3.5.2 Mobile Application

In order to offer an alternative and more compelling way to participate in the EIX, we decided to implement a mobile trading application (EIX-Market-App hereafter EMA). The implementation consists of two parts; a back- and frontend. In a first step we developed an application programming interface (API) to EIX's backend. We choose to implement a SOAP-interface, since it's interoperable and language-independent. The EIX-SOAP-API consists of nine methods related to trading and two methods needed to track user behavior. A WSDL file was used to generate the needed code with WSDL2Obj-C, an open source code generator. As the core data model of EIX and EMA are slightly different, objects containing the intersection of both models are used to exchange data between EMA and EIX. Second and regarding the front-end, we developed the EIX-Market-App as a mobile iPhone client for the EIX. EMA contains all core features of the EIX, i.e. submit and cancel orders, check the own holdings in the stocks-depot, access additional information like the order book, news, etc (Figure A.2). EMAs frontend-design is a compromise of two design goals. First, EMA was intended to be easy to use for new users. Second, existing EIX-users should be able to use the App with minimal learning-effort. Due to the limited screen size of the iPhone platform, it is not reasonable to use EIX's web-interface without adaptations. EMA's frontend tries to be close to EIX's web-interface by sharing the same menu-structure and nomenclature. Analogous to EIX's web-interface, EMA offers seven stock related information screens linked from the trade screen (Figure 3.3). To allow research about user's information usage prior to submitting an order, the consumption of the six information panels (respectively screens, in case of EMA) are logged in both systems. EIX and EMA both track the time a user needs to create and submit an order as well as the information used in this process (Teschner et al., 2012).

3.6 Summary

A wide and important range of policy decisions are made on the informational basis of economic forecasts such as Inflation. However, it is a well established fact that traditional economic forecast models lack the necessary accuracy (McNees, 1992; Schuh, 2001; Osterloh, 2008). Simplified, the current approaches mix expert knowledge with historic extrapolation. They are thus inadequate to capture rapid economic changes.

We described a prediction market specifically designed to forecast the most important economic figures in Germany up to a three quarters in advance. The market design tries to circumvent some known drawbacks which came up in previously employed digital contract markets. Moreover, we drastically increased the potential forecast horizon from a couple of days to a couple of months. Besides testing a new market design in the field, the EIX-market offers a good platform to run web-based experiments. In the following chapter we will start by evaluating the market according to the measures described in the second chapter. Thereafter we will present results from three experiments which were conducted within the EIX.

Chapter 4

EIX - Data & Market Statistics

4.1 Introduction

IN the following section we apply the presented prediction market (quality) framework on the EIX prediction market. By splitting the market period in two phases, the first and the second round (year or version) we try to draw conclusions for market engineering. We start by presenting some descriptive market statistics and then evaluate the market design according to the previously described framework. We will show that (1) the EIX market is an active liquid market with (2) low and improving forecast errors and (3) performs well in comparison to the Bloomberg survey.

4.2 Market Activity

The following data includes the time span from 30th October 2009 till 31st of October 2011. In total 1,235 (1,006 in the first round) participants registered at the EIX market, of those 809 (680) submitted at least one order. Altogether participants submitted 79,334 (45,808) orders resulting in 34,028 (22,574) executed transactions. In the respected time frame 107 (47) stocks were paid out. On average every stock was traded 645 times (SD. 424). Figure 4.1 depicts the number of orders per day as a proxy for market activity over time. Due to a novelty effect the activity was quite high in the first month. Activity went down to a stable level and remained their for the rest of the experiment period. In general, one

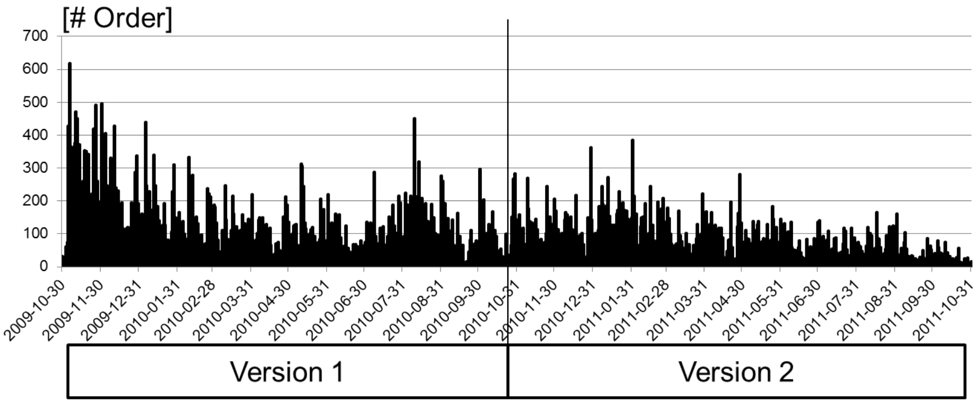

Figure 4.1: Market activity (number of orders per day) over time

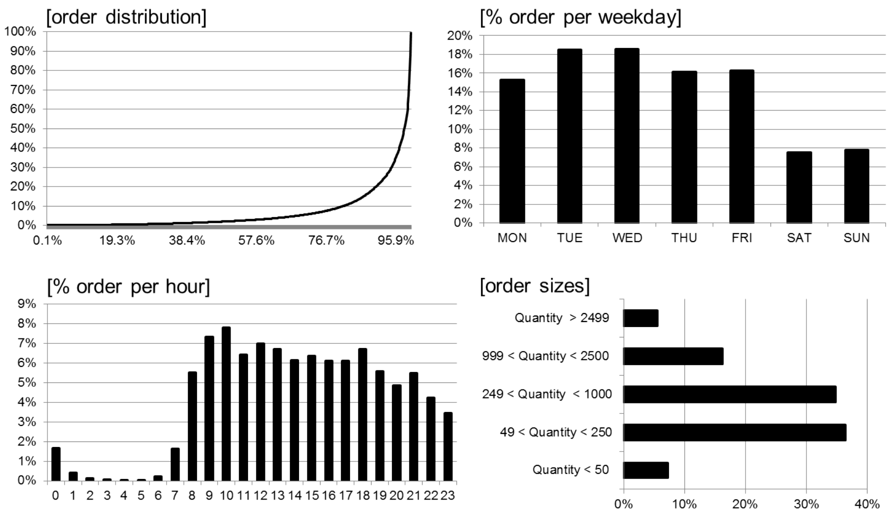

Figure 4.2: Market activity from different perspectives

can see that trading activity was lower in the second version (on average 127 vs 86 orders per day.). One reason might be the lower overall prize value that participants could win. The market activity, quantified as the number of orders submitted, is displayed in four dimensions in Figure 4.2. The number of orders per user is power law distributed. A few (power) traders submit the majority of orders (Gini inequality coefficient of 0.86). As the market is open 24/7 whereas financial markets operate only during office hours the trading activity is spread out over the day. Trading activity is quite evenly distributed between 6 am and 1 am o'clock, with slight peaks during noon. There is high activity during work-

Order Type	Number of Orders			
	Round 1 (V1)		Round 2 (V2)	
Buy	22,776	49.72 %	17,030	50.80 %
Sell	23,032	50.28 %	16,496	49.20 %
Total	45,808		33,526	

Table 4.1: Order statistics split by round and type

ing hours, and higher activity during the week compared to the weekend. It seems that people trade while at work. On average users submitted orders with an order size of 759 units (median 300). With prices between 90 and 140 this results in a volume turnover of 93,320 (median 38,690). In Figure 4.2 the number of submitted orders in each trade size category is displayed. Over 70 % of all orders are sized between 50 and 1000 stock units.
We expected to find activity to be unevenly distributed between indicators, as some arguably receive more media attention than others (e.g. GDP vs. Investments). But looking at the different indicators, our analysis shows that trading activity is evenly distributed among them (Figure A.6). However, trading activity within one stock increases as the liquidation date approaches. Hence, trading activity is the highest right before the stock is paidout. Splitting the number of orders according to their type we see that users submitted almost as many buy as sell orders (Table 4.1). The differences are insignificant in version one and significant in version two (diff: 1.6%; $t - stat. : 2.9; p - value < 5\%$).

4.3 Market Liquidity

The essential characteristic of a liquid market is that there are ready and willing buyers and sellers at all times. Hence, the usually used liquidity measures are adapted to prediction markets. Realized spreads and price impacts for example are usually denoted in 15 or even 5 minute intervals. We calculate the realized spreads and price impacts for longer time intervals (3 to 24 hours).

Table 4.2, first row presents the spread measures. The difference between

	Version 1			Version 2		
	Mean	Median	SD	Mean	Median	SD
Quoted Spread	171	78	367	67	38	171
Effective Spread	168	72	462	63	32	330
Quoted Spread at Trade	146	63	428	58	29	329
Realized Spread 3h	87	4	1043	22	19	247
Realized Spread 6h	80	4	1070	28	19	255
Realized Spread 12h	80	4.4	1055	25	19	153
Realized Spread 24h	78	4.4	1038	24	19	127
Price Impact 3h	71	15	1024	30	4	244
Price Impact 6h	78	17	1079	24	4	245
Price Impact 12h	78	18	1073	27	5	146
Price Impact 24h	81	2	1070	28	7	118

Table 4.2: Basic liquidity measures, averages over all stocks

quoted spread and quoted spread at trade (25 bps; $t - stat.: 12.3$; $p - value < 0.1\%$ in version one) shows that market participants observe the market and only actively trigger transactions by submitting market orders when spreads and thus implicit trading costs are low. Since quoted spreads at trade only measure the trading costs for the smallest of trade sizes, a more accurate measure of execution costs are given with the effective spreads. The realized spread represents the part of the effective spread that a liquidity supplier keeps as revenue. The price impact measures the information content of a trade. It reflects the permanent impact of a trade under the assumption that information impacts are permanent (Harris, 2002). As Table 4.2 shows, the price impact increases if the measurement time is longer and the realized spread decreases. One can follow that the market needs some time to adopt to the information brought in by trades. Furthermore by comparing spreads between version one and version two, we find that the liquidity significantly increases in the second version. For instance the quoted spread is 104 bps ($t - stat.: 66.2$; $p - value < 0.1\%$) lower in the second version.

Table 4.3 displays the average quoted spreads separately for the five different indicators. The last column gives the out-of-sample historic variability[1] of

[1] The sample ranges from 2005 until mid 2009.

	Quoted Spread	hist. Variability	Illiq	CSS
Exports	273	7.7	3.798	0.009
GDP	133	1.9	0.416	0.004
Ifo	231	11.7	0.453	0.012
Inflation	91	0.7	0.359	0.003
Investments	395	11.7	2.727	0.006
Unemployment	50	1.7	0.383	0.001

Table 4.3: Quoted spread, historic variability and spread estimates. Average values per indicator (Version 1)

each indicator. The variability can be interpreted as a risk measure for investors. Participants recognize the underlying risk as highly variable indicators, such as Investments and Exports, have high spreads and indicators with low historic variability exhibit low quoted spreads.

$$Quoted\ Spread = i + \beta * Variability \tag{4.1}$$

Running an OLS regression on a quote by quote basis the estimate (β) is 13.5, ($t-stat.: 15.5$; $p-value :< 0.1\%$). An increase in the variability of 1 point increases the quoted spread of the representing stock by 13 basis points on average. This seems reasonable as the market participants acknowledge the underlying high uncertainty and set the spreads accordingly.

In many prediction market datasets (e.g. Intrade) the orderbook data is not available for analysis. However, in order to judge market quality it would be beneficial to take liquidity into account. Hence, one might want to estimate the spread. In order to be able to use such estimates, we must first evaluate if spread-estimation measures can be applied in (thin) prediction market settings. As we have the actual spreads available in the EIX market, we can correlate estimates and spreads. The exact definitions of the measures are given in chapter two. Next, we will apply and validate the spread estimates. Table 4.3 depicts the average estimate for each indicator separately. For comparison reasons we also present the quoted spread. We correlate the different measures on an indicator level using following OLS regressions:

$$Quoted\ Spread_{i,t} = \alpha + \beta * Illiq_{i,t} + \sum_{j=1}^{5} \delta_j M_j \tag{4.2}$$

$$Quoted\ Spread_{i,t} = \alpha + \beta * CSS_{i,t} + \sum_{j=1}^{5} \delta_j M_j \tag{4.3}$$

In order to control for indicator effects we add the indicator dummies $M_1 - M_5$. The CSS measure is positively correlated ($t - stat.: 33.1$; $p - value < 0.1\%$), the Illiq measure is negatively correlated to the quoted spread ($t - stat.: -4.4$; $p - value < 0.1\%$). As expected, the CSS measures liquidity and Illiq measures the market illiquity. We find that both measures can be used to approximate liquidity in thin prediction markets.

4.4 Information Measures

In prediction markets in which we can observe the outcome, i.e. the fundamental value of each stock we can ex-post measure the information content of each order. This can be used to identify well informed traders, so called lead-users (Spann and Skiera, 2004). For prediction markets in which we cannot readily observe the outcome such as the foresight exchange (Graefe et al., 2010) or innovation markets (Stathel et al., 2010) one might use the price impact as a proxy for the information content of a trade. The price impact approximates the permanent information impact of a trade. As the price impact of each order can be measured at various points of time after trade execution, we evaluate four different time points and correlate the results on a trade by trade basis with the *EIX-Score* (equation 2.8). As one can see from Table 4.4 the later the measurement point (PI_{24h}) the higher the correlation with the *EIX-Score*.

We can conclude that such a measure can successfully be applied to identify well informed traders, even though the outcome might not be known. A more detailed analysis of identifying valuable participant input and experts can be found in Teschner and Weinhardt (2012a).

	PI_{3h}	PI_{6h}	PI_{12h}	PI_{24h}
Score	0.012	0.015	0.023	0.039
p-value	$< 10\%$	$< 5\%$	$< 1\%$	$< 0.1\%$

Table 4.4: Correlating: Score and price impact at different time points

From a market operator perspective it might be interesting to evaluate the probability of informed trading (PIN, see equation 2.10). For example might be to measure the effect of market micro-structure changes on the probability of informed trading. In our case we measure the effect of days till stock liquidation on PIN. The reduction of PIN is illustrated in Figure 4.3. Regressing the number

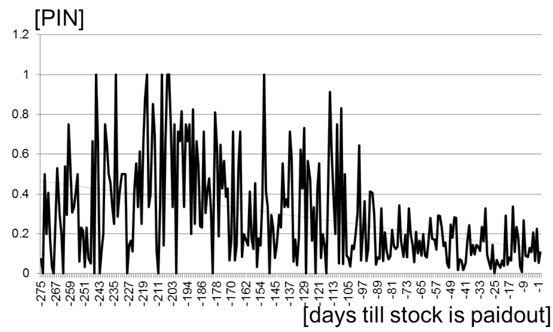

Figure 4.3: Probability of informed trading. (Number of days before an instrument is paid out)

of days before an instrument is paid out on PIN, we find that the probability decreases before instrument liquidation (estimate: -0.001 per day; $t - stat.: 4.19$; $p - value < 0.1\%$). In the regression we additionally controll for the absolute number of trades which has no influence on PIN. This is in line with finance literature, which suggests that PIN measures the risk that a trader faces a better informed trader. This also implies that uncertainty about the underlying event is reduced over time.

4.5 Forecast Performance

As previously described, forecast performance is a multi-dimensional concept. Hence we will describe the results from various perspectives. We aim at externally validating the generated forecasts by comparing them to the Bloomberg survey forecast, the industry standard. Furthermore, we will internally validate different contract designs by comparing forecast errors between EIX version one and EIX version two. We start by detailing simple error measures.

4.5.1 Error per Indicator

On an aggregated level we compare the market generated forecasts eight days before the data release ($forecast_{8,i}$) to the fundamental value. Table 4.5 summarizes the findings. We start by testing for a forecast bias. Hence, we test if the mean of the market forecasts is different from zero. As the t-statistics indicate there is no systematic bias in the forecast. This holds when separating the first an second version. As the mean error in the second version seem to be smaller we test on the difference between the two versions as well. However there is no significant difference. Testing for the difference in the standard deviations we find that the version two has a lower variance ($F - stat.: 33.6, p < 0.1\%$).

	Pooled	Version 1	Version 2	Difference (V1-V2)
Mean Error	1.76	3.47	0.33	3.14
SD	16.8	24.5	4.24	16.8
Bias (t-stat.)	1.1	0.9	0.6	-

Table 4.5: Forecast performance comparison of version one and two. Average price difference (error) between fundamental values and stock prices t-8. Testing for a systematic bias (if the mean error equals 0)

By comparing the standard deviations between forecasts and fundamental values we find that the produced forecasts are significantly less volatile than the fundamental values (1.41 vs 2.92; $F - stat.: 4.29, p < 0.1\%$) in the respected period. Thus we conclude that market forecast are more stable than the outcome values. This is in line with forecasts from other methods (Vajna, 1977). A reason for this is that forecasters regularly tend to publish moderate, conservative estimates rather than extreme values.

Table 4.6 reports various error measures for the different market categories. In order to make the market forecasts comparable, we transform the market prices back to predictive values. The transformation is the inverse payout function defined in Table 3.1. We compare the forecast errors (a) within the market and (b) between the market and two external benchmarks. The first simple benchmark is a naive forecast which can be created by using the last value to predict the following. Usually this is referred to as Autoregressive Model (AR(1)-Model). The other benchmark is the Bloomberg consensus forecast. Consensus forecasts has provided macroeconomic forecasts for industrialized countries since October 1989. Every Friday consensus publishes a number of prominent financial and economic analysts, and reports their individual forecasts as well as simple statistics summarizing the distribution of forecasts. For comparison reasons we use the mean of the forecast distribution. It is also this *"consensus"* forecast that receives the most attention as a summary assessment of the views of the private sector (Prakash and Loungani, 2001). As the publication date is fixed on Fridays the time between the forecast and the official data release varies. In order to ensure that the market forecast is unaffected from the Bloomberg forecast we use the market generated forecasts eight days before the data release. Forecasts are provided for all but for the Investment indicator. As the indicators have different baselines and units, the absolute error (MAE, Table 4.6) does not allow a comparison between various indicators but only a comparison to Bloomberg. The same holds for the MAPE. It becomes clear that some categories exhibit higher error values than other. This stems from the fact that some indicators such as Investments and Exports are highly variable. Moreover, the difference in the MAPE between the first version and the second version is due to different baselines. The Exports in version one are measured as the percentage change (MoM), whereas in version two the market predicts the absolute number in billions. In order to measure the relative accuracy within the market and between different categories we normalize the error values using Theil's U. The higher the Theil's U the lower the relative accuracy (Leitch and Tanner, 1991). Moreover, if the Theil's U statistic is lower than one, the forecasting technique is better than the naive forecast. As all Theil's U values are lower than one, we conclude that on average the market beats a naive forecast. To put that in perspective, Osterloh (2008) shows the naive forecast is often as good as the expert prediction for economic indicators. We see that exports are better in round one than the Bloomberg forecast, however in version two they perform equally well. The same is true for the GDP forecast. In the Inflation and Ifo-Index the Bloomberg forecast performs better. Finally, the market outperforms Bloomberg in the unemployment indicator. Hence, we find that the direct forecast comparison shows that they perform equally good.

Next we will analyze a within-market measure; the information efficiency. Table 4.7 presents the results of the previously presented OLS regression testing for forecast efficiency (see regression 2.13). If one of the estimates would be economically significant different from zero, the forecast revisions would not follow a random walk and one could assume a bias. However we find significant negative autocorrelation. Hence, the weak form forecast efficiency is not fulfilled which means that the forecast at a certain point of time does not contain all available information. Moreover, we find both the coefficients and the t-statistics to be higher in round two. It seems that information aggregation worked better in the first version.

We also run a Fair and Shiller (1989, 1990) regression, attempting to predict economic data releases on the basis of our two alternative forecasts. In order to compare both forecasts, we use a market forecast from eight days before the data release, which in all cases is before the Bloomberg release. Comparing the forecasts; the Bloomberg-based forecast has a large and extremely statistically significant weight, describing that it encompasses all information. Investigating further we find that the Bloomberg's superiority is mainly due to the market's poor performance in the exports indicator. Comparing the two version, we find that the market carries more information in comparison to Bloomberg in the second round. There are two possible reasons, the first is that Bloomberg's forecast

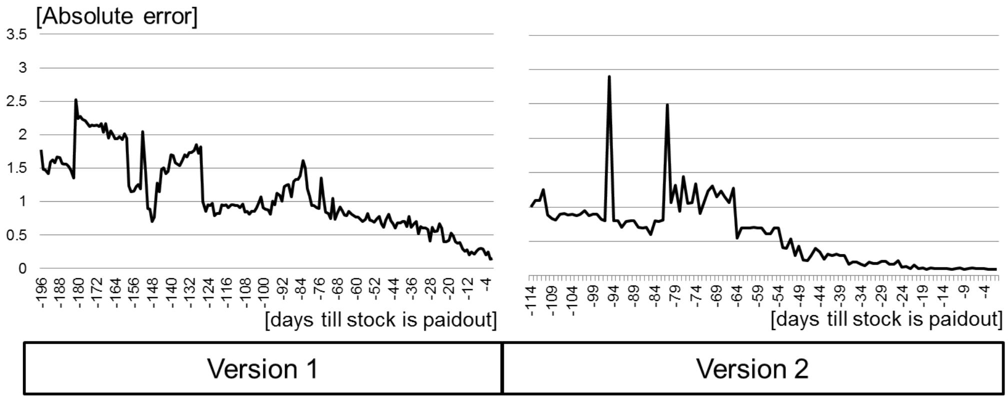

Figure 4.4: Absolute forecast error over time. Left side first version vs. second version right side.

performance decreased over time. This seems unlikely given there long experience and unchanged methodology. A second explanation could be that the market performance improved over time. Taking the results from Table 4.6 and Table 4.5 into account, the second explanation seems more likely.

4.5.2 Forecast Error Reduction

An important question is if the market continuously aggregates information. In Figure 4.4 the average absolute error over time is depicted. One can see a steady decreasing absolute error in the last 70 days.

$$AE = i + \alpha days + \sum_{j=1}^{5} \delta_j M_j \tag{4.4}$$

We run a OLS-regression analysis to quantify the error reduction per day. In order to control for different indicator effects we add the market dummies $M_1 - M_5$. We then apply the same regression for different markets and the two versions separately. Table 4.9 presents the results.

In the last 70 days the average error is reduced by 0.012 per day ($t - stat.: 7.0$; $p - value \leq 0.1\%$). Turning to the differences between version one and two, we see that the coeffient is lower in the second version. This due to the fact that in version one the error on day 70 is much higher than in version two (see Figure

4.4). Hence the reduction per day is lower in the second version. We conclude that forecast uncertainty was reduced over time, information aggregation took place and hence the absolute error was reduced.

4.5.3 Predicting Forecast Error

Another important question for interpreting point forecasts is the uncertainty attached to the forecast. Neither Bloomberg nor market forecasts provide explicit uncertainty information. However, one might interpret implicit market properties as proxies for the underlying uncertainty. One possible implicit market measure for market confidence is liquidity. Another proxy for market uncertainty could be a high price variability which indicates the traders' disagreement with the fundamental price of an asset. Furthermore, we include the difference between the highest and lowest price ($Distance_i$). Last but not least, the higher the number of traders who are active in one stock, the more likely it is that all available information has been incorporated (Van Bruggen et al., 2010).

$$AE_i = i + \sum_{k=1}^{5} \beta_k Predictor_{k,i} + \sum_{j=1}^{5} \delta_j M_j \tag{4.5}$$

We run an OLS regression to analyze if the four factors predict the forecast error magnitude. Table 4.10 presents the results. An increase in the quoted spreads by one point increases the forecast error by 1.7 points on average. This seems reasonable as the market participants acknowledge the underlying high uncertainty and set the spreads accordingly. All other implicit market measures have no explanatory value. As we included market dummies a difference in quoted spreads does not only indicate uncertainty due to differences in the underlying indicator variability but differences within one indicator.

4.5.4 Combining Forecasts

Combining forecasts can reduce error in several ways. A combined forecast is likely to be more accurate than a typical forecast of an individual component, because biases associated with the data and methods used in various forecasts are likely to differ.

For our study we combine the market generated forecasts with the Bloomberg forecast using an unweighted average.

$$CF_i = \frac{Bloomberg_i + EIX - Market_{8,i}}{2} \tag{4.6}$$

In order to test whether forecasts can be further improved by simply combining them, we compare the forecast errors. Table 4.11 lists the results. The forecast error is the lowest for the combined forecasts (testing the difference to Bloomberg; $t - stat.1.67, p - value < 10\%$). Moreover, the variance in the error is the lowest as well.

4.6 Summary

In this section we provided an in-depth analysis of a new market for economic outcomes. The following section highlights the findings from three perspectives.
Market design perspective: We first summarized findings from previous markets in this domain and detailed the known shortcomings of the currently used binary market designs. We proposed a radically different approach using a linear payout function. The theoretical improvements are threefold; first of all, the number of traded stocks is reduced. This leads to higher liquidity in the traded stocks. Secondly the "partition-dependence" bias can been avoided and lastly information can be aggregated continuously and over longer time horizons. Using the continuous market engineering approach, we tried to improve the second version by adapting the market. We find that in most measures the second version performs better than the first version even though the lottery-prizes and overall participation were lower. We strongly believe that by rigorously analyzing market properties there is potential for further improvement.
Forecasting perspective: The market acts as a mechanism not only to aggregate dispersed information but also to aggregate individual forecasts. It does so by incentivizing participation and rewarding early, precise forecasts. Moreover the

EIX-platform is yet alone in aggregating these forecasts continuously and for a long time horizon. Turning to the community generated forecasts we find that forecast accuracy improves constantly over time and that generated forecasts performed well in comparison to the Bloomberg-survey forecasts. Furthermore, forecast for all indicators beat the naive benchmark forecast. Additionally, we are able to show that the market has three supplementary benefits. Market measures can be used to identify valuable user input and forecast experts in real-time. Detecting such input might possibly enable us to improve the information aggregation mechanism and the forecast performance of such systems. Additionally, we can show that the market measures (quoted spreads) can be used to predict the forecast error. Hence, the market provides an implicit but visible measure for forecast confidence. In future this might enables us to enhance the market forecast by providing this measure explicitly as forecast confidence. Finally, in line with previous work, forecasts can be improved even more by combining Bloomberg and market generated forecasts.

Market quality perspective: We rigorously applied the previously described framework. The defined measures allow us to compare different market designs on a meta-level in future work. Moreover, it allows us to benchmark against markets in which not the full dataset is available (e.g. Betfair or Intrade). Hopefully the evaluation of these spread estimation measures allow us to address the still missing prediction market meta study.

It seems fair to conclude that overall the market worked remarkably well. We hope our approach will positively impact the market design community and forecast results will eventually influence economic policy making in Germany by providing continuous information about the state of the economy.

Indikator	Unit	MAE		MAPE [%]		RMSE		Theils U		Sharpe Ratio	
		EIX	Bloomberg (BB)	EIX	BB	EIX	BB	EIX	BB	EIX	BB
Export V1	%	3.56	2.88	123.27	69.48	4.79	3.73	0.41	0.52	0.68	0.66
Export V2	Bil.	2.09	2.10	2.39	2.39	2.79	2.76	0.96	0.94	0.20	0.18
GDP	%	0.44	0.29	124.13	78.76	0.58	0.40	0.39	0.42	0.72	1.25
Inflation	%	0.18	0.11	21.91	12.31	0.27	0.16	0.74	0.39	0.09	0.53
Ifo V2	Points	1.34	0.89	1.21	0.81	1.71	1.06	0.81	0.57	0.27	0.43
Investments	%	1.21	-	121.40	-	1.44	-	0.50	-	0.99	-
Unemployment	Num.	37,750	89,792	1.16	2.80	51,922	12,3941	0.38	0.95	0.61	0.51

Table 4.6: Market forecast errors, compared within the market and benchmarked against the Bloomberg (BB) survey forecast

	rev_{t-1} (t-stat.)	rev_{t-2} (t-stat.)	rev_{t-3} (t-stat.)
Version 1	-0.03 (-5.2)	-0.02 (-3.8)	-0.01 (-3.2)
Version 2	-0.5 (-28.5)	-0.23 (-12.4)	-0.05 (-3.9)

Table 4.7: Weak-form forecast efficiency

	Bloomberg (t-stat.)	EIX (t-8) (t-stat.)
Exports	1.15 (2.5)	-0.13 (-0.3)
GDP	5.17 (3.8)	-4.14 (-3.2)
Ifo	1.15 (5.4)	-0.14 (-0.7)
Inflation	1.23 (6.1)	-0.23 (-1.15)
Unemployment	-0.02 (-0.19)	1.02 (9.9)
Version 1	1.39 (4.7)	-0.37 (-1.3)
Version 2	-2.6 (-1.7)	3.5 (2.4)
Pooled	1.25 (5.6)	-0.25 (-1.17)

Table 4.8: Fair-Shiller regression

	Version 1 (t-stat.)	Version 2 (t-stat.)	Pooled (t-stat.)
Export	-0.25 (-3.7)	0.01 (0.5)	-0.49 (-8.9)
GDP	0.12 (1.1)	-0.04 (-4.7)	0.05 (0.7)
Ifo	-0.05 (-1.2)	-0.07 (-11.9)	-0.06 (-6.3)
Inflation	-0.06 (-9.9)	-0.05 (-10.4)	-0.07 (-15.3)
Investment	0.11 (0.9)	-0.24 (-8.4)	-0.03 (-0.5)
Unemployment	-0.04 (-13.9)	-0.03 (-13.5)	-0.04 (-20.4)
Pooled	-0.08 (-16.3)	-0.05 (-1.8)	-0.12 (-7.0)

Table 4.9: Forecast error reduction per day (last 70 days)

Variable	Estimate	(t-stat.)
$Intercept$	-1.67	(-0.5)
$Quoted\ Spread_i$	1.77	(2.1)
$Var(price_i)$	-0.00	(-0.1)
$unique\ traders_i$	0.03	(1.3)
$Distance_i$	0.01	(0.1)

Table 4.10: Predicting market forecast errors. (Regression includes control dummies for the indicator categories)

	MAE (SD)	MAPE (SD)
EIX	11.50 (28.0)	0.08 (0.3)
Bloomberg	7.08 (20.3)	0.06 (0.3)
Combined	6.60 (19.3)	0.05 (0.2)

Table 4.11: Combining forecasts

Chapter 5

Short Selling in Prediction Markets

5.1 Introduction

SHORTLY after the beginning of the global financial crisis, several stock exchange regulators banned short selling on financial markets (SEC, 2008). They argue that this was necessary to maintain properly operating markets for securities trading and to ensure the stability of the financial system. An ongoing debate in finance is whether short selling has positive or negative effects on market quality. While financial market experts stress the importance of short sales for market liquidity and price discovery, many retail investors attribute the market turmoil in 2009 also to short selling. Analyzing the EIX prediction market data, we assesses the consequences of the introduction of short selling on market forecast accuracy, a proxy for market quality. Using an event-study approach we find that introducing short selling further improves the EIX market forecast accuracy. By allowing traders to short sell, mispricing is reduced and hence market forecasts are closer to actual macroeconomic outcomes.
The remainder of this chapter is structured as follows: the next section starts by summarizing previous work on short selling effects on market quality and efficiency. Additionally we present the evaluation methodology. The subsequent section then analyzes the market from a forecasting perspective and quantifies the effect of short selling. Finally, section four concludes this chapter.

5.2 Related Work

We believe that the concept of short selling is important to complex prediction markets and facilitates predictive power. But what changes for participants if they are able to short sell in a prediction market? The calculation of a stock's value on the buyer side is quite simple. If a trader estimates a stock's expected future value higher than the best price asked by a seller, then this trader should buy this stock. The calculation for sellers is a bit more subtle with short selling. Traders who believe that a stock's future value is lower than the best current price offered by a buyer should sell short, i.e. accepting the current bid price in exchange for a liability that is based on a stock's future value.

5.2.1 Short Selling in Financial Markets

In financial markets there is an important distinction between covered short sales and naked short sales. Covered short selling occurs if an investor actually possesses the shares that she is selling, for instance through borrowing shares from a third party. The investor then usually pays a lending fee to the owner of the shares. When such a short position is closed, the investor returns the shares to the lender. In financial markets a third party lending shares might be a bank or insurance company with a large portfolio of securities. In the case of naked short selling a trader does not possess the shares she is selling and she does not make any arrangements to borrow shares at the time of the sale. Often, only market makers are allowed to short sell naked. They resolve their inventory intra-day or at least before a trade is settled (Boulton and Braga-Alves, 2010).
Financial research suggests that short selling is an important mechanism for efficient prices and the reflection of private information through asset prices. Jones and Lamont (2002) find support for the hypothesis that stocks which are expensive or impossible to short are overpriced. In trading environments with little public information short sales are important to reveal private information through order flow (Cohen et al., 2007). Short sales in general disclose bad news to markets (Aitken et al., 1998). Additionally, short sales can improve price accuracy. On the one hand they give incentives to traders to gather new information,

on the other hand they help to better reflect already available information in the market (Fox et al., 2011).

5.2.2 Event Studies

Event study methodology is a widely used instrument in financial market research. It enables the researcher to capture the effect of an event on a share price. In order to do this, a time series of the asset's return is needed, e.g. on monthly or daily basis. The time series is split into an estimation window and an event window. According to equation 5.1 an abnormal return AR_{it} for asset i at time t is computed. R_{it} denotes the actual return and K_{it} the expected return, also called normal return (Brown and Warner, 1980).

$$AR_{i,t} = R_{i,t} - K_{i,t} \tag{5.1}$$

The simplest model for estimating normal returns is the so called constant mean return model (CMR) (MacKinlay, 1997). It establishes the normal returns as the mean of the returns in the estimation window. Other models use the correlation with market wide return movement in order to calculate the normal return. Finally, abnormal returns are aggregated over all examined assets and are tested for significance in the event window.

5.2.3 Event Studies in Prediction Markets

Event studies were already successfully used on prediction markets. Elberse (2007) conducts an event study on the Hollywood Stock Exchange and finds that star actors have a positive impact on movies' revenues. Slamka et al. (2008) compare the performance of the event study methodology in a play-money and in a real-money prediction market. They find that both settings are likewise suitable for conducting event studies.

In order to capture the impact of short selling on forecast accuracy we use an event study approach on daily basis. We choose to set the estimation window to 30 days. The event window is the day where short sales are introduced. At the end of each day t in estimation and event window we calculate the $error_{i,t}$ for asset i as follows:

$$error_{i,t} = |fv_i - Mid_{i,t}| \tag{5.2}$$

In the equation fv_i denotes the fundamental value and $Mid_{i,t}$ the mid price. For the event study methodology the relative error change can then be interpreted as a return. Abnormal and normal relative error changes can be interpreted analogously. We check for occurrences of potentially disturbing events in the event window with the help of the economic news-stream of the EIX market, but do not find anything striking.

Following Elberse (2007) we use a constant mean return model to estimate the normal relative error change. We test the significance of the abnormal relative error change with the t-statistic proposed by (Brown and Warner, 1985). The test statistic is calculated as follows:

$$t - stat. = \frac{\overline{error}_t}{\sqrt{Var(\overline{error}_t)}} \tag{5.3}$$

Where $\overline{error}_t$ is the mean of errors at event time t. The test statistic converges to a standard normal distribution.

Additionally, we conduct a nonparametric Corrado (1989) rank test in order to verify the results without assuming underlying standard normal distributions. For the Corrado test, ranks are assigned to the abnormal returns in ascending order. Let $t = 0$ denote the event day, $\kappa_{i,0}$ the rank of the event day of asset i and $E(\kappa_i)$ the mean of the assigned ranks. Then the test statistic is computed according to equation 5.4.

$$t - stat. = \frac{\frac{1}{N}\sum_{i=1}^{N}(\kappa_{i,0} - E(\kappa_i))}{\sqrt{Var(\kappa)}} \tag{5.4}$$

Where $Var(\kappa)$ is the variance of $\frac{1}{N}\sum_{i=1}^{N}(\kappa_{i,t} - E(\kappa_i))$. The test statistic is asymptotically standard normally distributed.

5.3 Setting and Research Question

Introducing the concept of short selling in a complex prediction market allows us to test the effect of short-selling on market quality and forecast performance. The deviation between the market prediction and market outcome provides a direct test of market efficiency. Hence, the research question is whether the introduction of short selling reduces the deviation (e.g. forecast error).
In order to test the effect of short selling on market efficiency and quality with an event study approach, we gradually introduced short selling in the EIX market. This means that short selling was allowed in single stocks from March 15th till the 31st of October 2010. We introduced short selling for each stock separately. Short sales were allowed at noon, 15 respectively 25 days before the data release. This allows us to analyze the effect for each individual stock.

5.4 Results

The following section first presents some descriptive market statistics and then evaluates the introduction of short sales according to the previously described methodology. We will show that short-selling improves the forecast accuracy and market liquidity.

5.4.1 Descriptive Statistics

The following data includes the timespan from 30th October 2009 till 31st of October 2010. In total 1,006 participants registered at the EIX market, of those 680 submitted at least one order. We discard all stocks with less than 50 transactions. In the respected time frame 47 stocks were paid out. In 29 stocks short selling was introduced.

5.4.2 Market Performance

We conducted an event study for each indicator and the whole market separately. Table 5.1 shows the results. Day 0 denotes the event day. A negative day

	Day 0 Rel. Abn. Error Change	t-Test p-Value	Corrado Rank Test p-Value	No. of assets
All indicators	-0.31	< 5 %	< 1 %	29
Exports	-0.04	n.s.	< 5 %	8
GDP	0.11	n.s.	n.s.	3
Inflation	-0.65	< 1 %	< 1 %	8
Investments	0.62	n.s.	n.s.	2
Unemployment	-0.62	< 10 %	< 1 %	8

Table 5.1: Results of the event study. (n.s.: not significant)

0 relative abnormal error change corresponds to an abnormal high forecast improvement when short selling is introduced.

Jointly for all indicators the t-test shows significance at the 5 % level for an abnormal error change of about -30.8 %. Forecast errors drop more than -60 % abnormally for Unemployment and Inflation asset categories. They are significant at the 10 % level, respectively 1% level. The Corrado rank test supports our findings.

The small abnormal error change of 3 % for Exports shows significance at the 5% level only for the Corrado rank test. A possible explanation for this is that Exports are highly volatile and therefore traders have difficulties to evaluate the Exports assets, even with short selling allowed.

Results for GDP and Investments are not reliable as the test statistics are poorly specified for only 3, respectively 2, data points. All in all, we draw the conclusion that introducing short sales improves forecast accuracy of prediction markets.

5.4.3 Market Liquidity

Another important question for interpreting point forecasts is the uncertainty attached to the forecast. Neither Bloomberg nor market forecasts provide explicit uncertainty information. However, one might interpret implicit market properties such as quoted spreads as proxies for the underlying uncertainty. Hence, we check, if quoted spreads change after introducing short selling. In order to do so, we compare the spreads 30 days prior to the event day to the spreads past

	Pre-event Mean	Post-Event Mean	Satterthwaite t-Test p-Value	F-Test p-Value
Exports	0.047	0.027	$< 0.1\%$	$< 1\%$
GDP	0.017	0.007	$< 0.1\%$	$< 0.1\%$
Inflation	0.013	0.005	$< 0.1\%$	$< 0.1\%$
Investments	0.036	0.015	$< 0.1\%$	$< 0.1\%$
Unemployment	0.004	0.002	$< 0.1\%$	$< 0.1\%$

Table 5.2: Market liquidity pre and post short selling. Satterthwaite t-Test and F-test results of quoted spread changes.

event day for each market segment.

Results are shown in table 5.2. For each market segment the F-test shows strong significance below the 1 % level. Therefore, the equal variance hypothesis must be rejected. We are aware of the fact that to some extent variance change of the spreads may be due to closeness to the data release and because of that uncertainty might be reduced. Nonetheless, there is a strongly significant change of spread variances after introduction of short sales. Furthermore, the differences of post- and pre-event sample means are striking. A Satterthwaite t-test is significant at the 0.1 % level for all market segments. We conclude that introducing short sales reduces quoted spreads respectively forecast uncertainty in prediction markets.

5.5 Conclusion

Based on existing financial research we believe that short sales in prediction markets are an important factor to improve prediction accuracy and reduce biases. This study supports this believe. In general, prediction markets function best if market participants have the same market power independent of trade direction. The decision to buy or sell a stock should mainly be driven by a trader's estimate of a stock's future value and not based on whether trading on expectations can be achieved through buying or selling a stock. Comparable to financial markets, short selling needs some restrictions in prediction markets. If traders want to buy stocks they are limited through their initial cash endowment and subsequent cash changes. Since short selling involves selling stocks that traders do not possess there is no natural corresponding limit for selling stocks. In short selling enabled prediction markets, there is the need for a short selling restriction equivalent to the budget constraint for buying stocks.

There are several options to implement short selling in prediction markets and to implement short selling constraints. The market operator could generally function as a credible lender of stocks. It is important to make sure that market participants are liquid enough to meet their short obligations once trading in a specific stock ends. One basic approach is to implement a fixed constraint through the number of shares that a trader can go short. Other possibilities include dynamic implementations which are based on potential losses a trader would have to realize in extreme cases through short selling. In our opinion there is no general rule how to implement short selling constraints. Such limits are rather prediction market specific.

In a first step we present one way how short selling can be implemented in a linear prediction market. In order to quantify the effect of the introduction, we choose to allow short selling separately for individually stocks. All in all we collected 29 time points when short selling was introduced. We then document the positive effect of short selling on market accuracy using an event study approach. Analyzing market quality, we find that due to the market changes, spreads which are a proxy for market uncertainty are lowered.

We conclude that short selling in linear prediction markets is an essential characteristic to enable high prediction accuracy. Future research involves empirical testing of different short selling implementations. It is important to analyze whether specific short selling restrictions are actually needed and how trader behavior changes through short selling.

Chapter 6

Incentives, Feedback & Learning in Prediction Markets

6.1 Introduction

INTERNET communities -such as the EIX-market- offer the advantage of instant information exchange and group decision that is not possible in a real-life. In the previous chapter we showed how such an online community can been used to facilitate information aggregation of macroeconomic variables. This subsequently leads to the question of how participants can be motivated to contribute and share their information for longer time horizons. We try this by first implementing a specifically designed market environment. Secondly, we design a play-money incentive schemes which rewards participants according to their performance. How well does this incentive scheme fulfill the goal of keeping participants active and contributing? It is especially interesting of how good this incentive system works for longer time horizons. An important part of the incentive mechanism is the communication of it. Participants have to intuitively understand how their actions relate to an outcome and finally to a performance-based (incentive) ranking. In (financial) markets this feedback loop is inherently part of the mechanism. Traders gain or lose money. However, as prediction markets usually rely on play-money, it seems worth considering designing a feedback loop linking the actions and outcomes.

In public goods projects, non-monetary and participation feedback has been found to increase participants' contributions. In order to further motivate participants intrinsically we test five different feedback mechanisms. We test which of the feedback types works best at motivating contributions. If we find any differences in the activity level, do the additional contributions improve the community forecasts?

Related to the question of feedback is the notion of learning. Classical learning-by-doing models suggest that traders might improve their ability as they actively trade. Through their actions and provided feedback traders gain experience and thus improve over time. The question arises how to improve the participant's learning process. In our platform learning occurs on three levels. Firstly, participants learn how to trade in a continuous double auction just as in stock markets. Secondly, they learn about their own macroeconomic forecasting ability in comparison to their peers. Thirdly, by following market forecasts they learn about the current state of the economy. Separating two distinct learning types we analyze how learning takes place within the participating community. Furthermore, we evaluate the effect of performance feedback mechanisms on activity in a market-based system.

The remainder of this chapter is structured as follows: The second section gives a brief review of incentives schemes and feedback in information exchanges. Furthermore learning in markets is discussed. Section three presents the research questions. The subsequent section evaluates the EIX-market from the three perspectives; incentives, feedback and learning. Finally section five concludes this chapter.

6.2 Related Work

6.2.1 Incentive Schemes for Play-Money Markets

As pointed out by Servan-Schreiber et al. (2004) there are two essential factors driving the prediction market accuracy; knowledge and motivation. One practical way to increase motivation is the promise of winning money. In real money markets traders invest money and gain directly like in financial markets. Due to the legal restrictions on gambling, setting up a real-money market incurs huge technical and regulatory costs. As an alternative, market operators can set up play money prediction markets. Instead of real money, participants are endowed with a virtual currency. Previous research has shown that play-money perform as well as real-money markets predicting future events (Wolfers and Zitzewitz, 2004; Rosenbloom and Notz, 2006). Gruca et al. (2008) state that there is no difference if there is a lot of publicly available information otherwise real money markets perform better. However, Rosenbloom and Notz (2006) argue that real money markets may better motivate information discovery while play money markets may yield more efficient information aggregation.

In order to encourage participation and information revelation in play money markets, market operators shuffle prizes according to designed incentives schemes. As there are various ways how such schemes can be designed, the questions arises if the design influences market performance and trader behavior. Luckner and Weinhardt (2007) study the impact of three different incentive schemes on prediction accuracy in short-term laboratory experiments. They find that that a rank-order scheme outperforms a fixed payment incentive scheme and surprisingly a performance-compatible payment. Their results show that rank-order tournaments are a suitable incentive schemes in case of risk-averse traders. Moreover, they argue that competition in the rank-order treatment overrides risk aversion and in doing so leads to the best results in terms of prediction accuracy.

Prediction markets work by incentivizing information revelation and participation. Hence traders can be rewarded based on their performance which is directly linked to the quality of their contributions. As Spann and Skiera (2003) point out that participant motivation decreases if payout dates are too far in future. It remains unclear how the participants can be incentivized in long term field prediction markets.

6.2.2 Feedback Mechanisms

IS literature indicates that system feedback highly influences the usage of information systems (Bajaj and Nidumolu, 1998; Kim and Malhotra, 2005).
A common way to increase participants' intrinsic motivation to contribute to public goods projects is to give users feedback and recognition. Cheshire and Antin (2008), as well as Ling et al. (2005), try to raise user contribution in online communities through feedback mechanisms. Many motivational theories in psychology include a feedback component such as the goal setting theory from Locke (2001). According to Cheshire and Antin (2008) there are three different feedback mechanisms which are assumed to lead to an increased contribution rate.

- *Gratitude* is a simple "Thank you"-message displayed after a contribution. Beenen et al. (2004) find that sending a one-time "Thank-you" email can raise contributions.
- A *Historical Reminder* is a feedback mechanism, which informs the participant about the number of individual contributions. According to Cheshire and Antin (2008) this may help the user to think about his own past contribution behavior.
- *Relative Ranking* displays the contribution frequency compared to peers. The knowledge about cumulative group behavior can be beneficial to the production of a public good, like contributions to a wiki (Cheshire, 2007).

Additionally one might consider to use *Social Ranking* feedback, which illustrates a ranking within a group of users who created a similar number of contributions. One might argue that information about individuals with a similar ranking, and therefore the same amount of contributions, leads to an increase in social competition, and positively impacts the motivation to contribute. However -to our knowledge- social rankings have not been tested yet. As most work on the effect of feedback mechanisms on user participation and contribution relates to cooperative environments such as public wikis, it remains unclear if and how the effects can be reproduced in more competitive environments such as electronic markets.

Besides the usage and adoption, feedback plays a major role in interacting with the IS-user. An assumption underlying current models of learning is that learning takes place only through repeated experience of outcomes (Weber, 2003).

6.2.3 Learning in Markets

As prediction markets work like financial markets, they offer a learning environment for trading in stock markets. Moreover, a recent study shows that prediction markets can enable active learning in large groups (Buckley et al., 2011). Participation in prediction markets changes the learning event from the passive receipt of material and recall of facts to active decision making. Thus learners are challenged to engage in the learning process. In a similar case-study a prediction market was used as a teaching tool for MBA classes (Raban and Geifman, 2009). The authors conclude that students gained valuable insight into their own decision making patterns as well as the hands-on activity helped to enhance the understanding of markets and added value to the lessons. However, both explorative studies use short-lived prediction markets as a pedagogical tool in closed class-room environments.

Classical *learning-by-doing* models suggest that traders might improve their ability as they actively trade. Through their actions and provided feedback traders gain experience and thus improve over time. A second type of learning is called *learning about ability*. As investors trade, they might realize that their ability is low and decide to stop trading. By analyzing investor records, Seru et al. (2010) separate these learning types and find that most of the learning occurs as individuals learn about their own ability and low-ability investors stop trading. Contrary to these results a study on retail investor behavior finds that excess portfolio returns improve with account tenure - a proxy for investor experience. Furthermore, they also find that trade quality significantly increases with experience (Nicolosi et al., 2009). In prediction market literature, learning has been viewed from a forecasting perspective. Based on models of information aggregation, Adams (2006) theoretically shows that when learning is allowed, a prediction markets may aggregate information. In particular, adding learning to the

Manski (2006)-model causes market prices to converge to the mean of the distribution of beliefs. Hence the *market* learns the correct outcome probability over time.

6.3 Setting and Research Question

The main research question is how participants can be motivated to contribute and share their information for longer time horizons. We design a play-money incentive scheme which rewards participants according to their performance (see Section 3.2.3). Accordingly we have to evaluate how well the incentive scheme fulfills the goal of keeping participants active and contributing. It is especially interesting to analyze how well this incentive system works for longer time horizons.

Another question is, if market participants can be intrinsically motivated by certain IS-artifacts. In order to test this we implement five common feedback mechanisms (see Section 3.3.3). We then evaluate which of the feedback types works best at motivating contributions. If we find any differences in the activity level, do the additional contributions improve the community forecast?

As discussed, feedback is essential for a successful learning process. This subsequently leads to question of how participants can learn about their contribution and improve their forecast performance over longer time horizons. Given that participants learn to improve their forecast ability, do market generated forecasts improve over time?

Extending the idea of a purely profit-based performance ranking we test if performance feedback can be improved. We display two types of ranking; one highlighting the overall trading performance, one more directly aimed at displaying forecast performance. How well do these artifacts work? Do participants learning from the performance feedback they receive in this setting? If we find any learning effect, does this improve the macroeconomic forecasts? This leads to the question of forecast accuracy in general. From a forecasting perspective, if learning about the outcome occurs, the flow of information reduces outcome uncertainty and hence results in decreasing forecast errors over time.

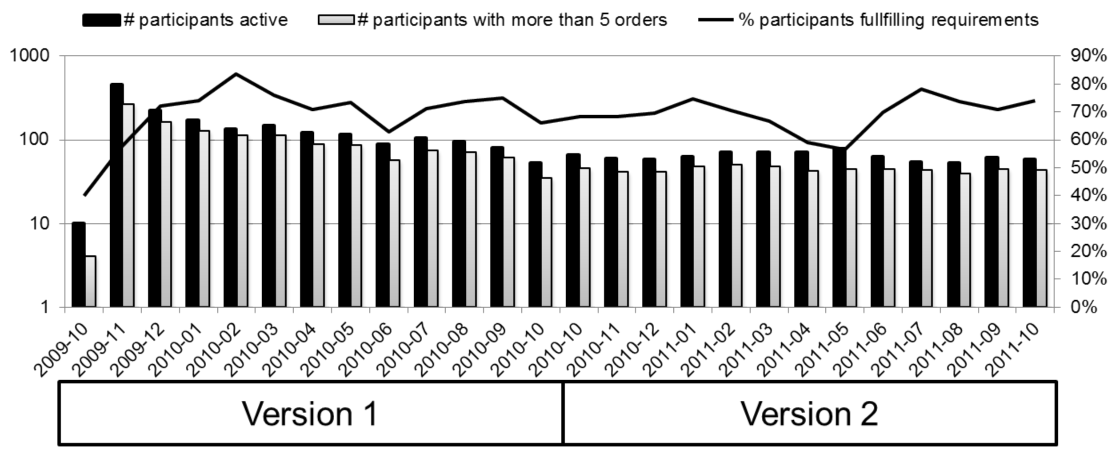

Figure 6.1: Participants staying active

6.4 Results

In the following section we analyze the three interlinked topics of incentives, feedback and learning.

6.4.1 Incentives

As described we designed an incentive scheme that aims at keeping participants' motivation high over the market run-time. Figure 6.1 presents the number of active participants on a monthly basis. We started the experiment with a small number of participants in the very end of October 2009. We started to promote the platform in the beginning of November 2009. This explains the low number of active participants in the first month.

Ignoring this kick-off effect, we find a clear novelty effect, which is evident in the high activity levels in the subsequent two months. While the number of active participants decreases we find that the percentage of participants fulfilling the monthly incentive requirements stays at the same level. The slight drop in the last month is due to the fact that we started the second market round in parallel. Over the second version, there is no general trend visible. Participants stay active and contribute continuously over the market run-time. We conclude that the incentive structure worked well for a such a long-running experiment.

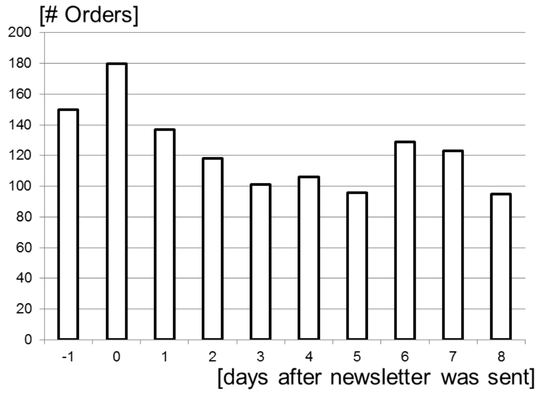

Figure 6.2: Effect of the newsletter

6.4.2 Feedback

In order to keep participants active and informed we sent out a weekly newsletter summarizing the recent economic news. The sending days varied during the week. Participants are able to sign-up (off) for receiving the newsletter on the EIX portal. At the end of round one, 63 % (round two: 58 %) of all active participants received the newsletter. Analyzing the impact of the newsletter, we find an increased activity measured as orders per day (on average +60 orders on sending days; $t-stat.: 3.23, p-value \leq 1\%$). The peak activity on sending days is followed in almost linear decreasing activity in subsequent five days (Figure 6.2).

We implemented five feedback mechanisms and added one control treatment (no feedback). On sign-up participants were randomly assigned to one of the treatment groups.

$$log(orders) = i + \beta * ML + \sum_{i=1}^{5} \alpha_i * F_i \tag{6.1}$$

We used the following OLS regression (equation 6.1) to test the influence of each feedback treatment on each individual activity level. In the baseline treatment no special feedback is given. As the number of orders is power-law distributed, following Raban (2008) we use a logarithmic transformation of the *Orders* variable. The regression results are depicted in Table 6.1. The model is dominated by

(N=410) log(Orders)	Newsletter	Thanks	Ranking	Trading performance	Number of trades	Random
Estimate	0.26^c	-0.01	-0.32	-0.22	-0.14	0.09
$(t-stat.)$	(1.9)	(-0.1)	(-1.2)	(-0.9)	(-0.6)	(0.4)

Table 6.1: Trading activity and feedback mechanisms: The superscript 'c' denotes significance at the 10% level.

the newsletter effect. Participants receiving the newsletter submit significantly more orders. As all feedback treatments show no significant effect, we conclude that feedback mechanisms do not induce any additional motivation to contribute in competitive market environments. It seems that different individual competitiveness levels dominate any feedback effect.

6.4.3 Learning

As previously detailed, learning might take place on two levels; first by actively trading, participants might gain experience and hence improve over time. Secondly, by observing their performance participants might realize their low-ability and consequently leave the market. Using the number of past orders as a proxy for experience we test the first idea of *learning by doing*. We run following OLS regression:

$$Profit_o = i + \beta_1 * NumOrders_{t,u} + \beta_2 * TD_o + \beta_2 * Init_o + \sum_{i=1}^{5} \gamma_i * M_i \tag{6.2}$$

$NumOrders_{t,u}$ denotes the number of orders a user u has submitted before the specific order o. We add five indicator variables (M_i) to control for indicator effects and two variables for trading behavior. TD_o indicates the trade direction of the order, it is one if it is a buy and zero for a sell order. $Init_o$ is one if the order is a trade-initializing order. We find the experience (numOrders) variable positively correlated (estimate: 1.62; $t-stat.: 3.51$; p-value < 1%) with profits.
Testing if participants learn about their own ability, we assume that participants are more likely to stop trading if their performance is below average. As men-

tioned we implemented and displayed two ranking versions, one displaying the overall portfolio value one aimed at showing the peer forecasting performance. When correlating the two rankings we find that they differ substantially ($\rho : 0.11, p - value < 5\%$). In order to test the *learning about ability* model we use performance data from the first round to predict activity in the second round.

$$Active_{u,2} = i + \beta_1 * active_{u,1} + \beta_2 * ML_u \tag{6.3}$$

The Logit regression (equation 6.3) tests if a participant was active (submitted at least one order) in the second round, dependent on being in the lower half of the portfolio ranking and was getting the newsletter. We find no significant effect for the overall portfolio ranking. However, if we recode the low-variable to reflect the forecast performance based ranking, we find that the below average performing participants are less likely to continue trading (odds-ratio: -1.95; χ^2:20.23; p-value < 0.1%). A reason for the difference might lie in the presentation, as the forecast-ranking is higher in the browsing menu structure on the webpage. Combining these results, we conclude that participants gain experience over time and with increasing experience submit more profitable orders. Turning to *learning about ability* we find that the forecast performance-based but not the portfolio-based ranking predicts if participants stay active.

6.5 Conclusion

Internet communities offer the advantage of instant information exchange and group decision that is not possible in a real-life. One important design aspect is the incentive mechanism that fosters participation and contribution. We present an incentives scheme well-suited to motivate participants contributing their information for longer time horizons. Investigating the level of participation, we find that activity is mainly driven by a weekly newsletter which acts as a reminder. Assuming that classical feedback mechanisms would lead to different participation levels, we find that the induced competitiveness of market environments seem to superpose classical feedback mechanisms.

Our semi-anonymous market enables naive and professional forecasters to test their forecast ability compared to their peers. We show that participants gain experience over time indicating that the active, engaging environment fosters learning. Furthermore, we find that learning takes place as participants who submitted more orders are more likely to submit an additional profitable order. Testing if participants are able to learn about their forecasting ability, we find that a specifically designed forecasting ranking provides the necessary feedback.

Chapter 7

Interface Design in Prediction Markets

7.1 Introduction

ONE reason for market failure is the inherent complexity excluding non-sophisticated users. The Internet has increased the number of complex (e.g. Energy, P2P resource sharing) markets dramatically. As more and more non-sophisticated users have to interact with complex markets, the question arises how to provide suitable interfaces for such users. Market complexity can be reduced by adopting the market rules or by simplifying the user interface. Just recently researchers started to address this topic and identified the need to merge market and interface design. More radically Seuken et al. (2010) proposed to hide most market complexities from the user. They call the this approach *Hidden Market Design*. The main idea is to hide or reduce market complexities while maintaining economic efficiency. One way to accomplish that is to simplify the market interface. However, it remains unclear how simplified trading interfaces affect market efficiency and individual trading decisions. Hence it is completely unclear how to actively design market interfaces.

A reason why non-sophisticated users might have difficulties interacting with such markets is the amount of information they have to cope with. A common belief about decision making is that the more information available the better are our decisions. In contrast to that intuitive belief, Human Computer Interaction (HCI) research has repeatedly shown that too much information leads to cognitive overload resulting in decreased decision performance (Eppler and Mengis, 2004; Chervany and Dickson, 1974). As a consequence we rely on information systems to filter, aggregate and present this information in a manner that supports the decision making process. Thus far there are no guidelines on what information is needed to support trading decisions. In order to create such guidelines, one needs to understand which information elements support and which elements hinder the individual trading process. Moreover, as individuals have different informational needs and vary in experience, it seems fruitful to develop customized market interfaces.

We propose to analyze user actions in a repeated market environment where information processing plays a key role. Within the EIX field experiment with more than 1,200 participants and over 65,000 single decisions we study the impact of information elements on trading behavior and performance.

In our prediction market for economic variables, participants have two ways to change their trading interface. First they are able to customize the default interface and second they can switch to a trading wizard. This provides us with datasets for two experiments.

In the first experiment traders can select up to seven information elements to adopt the interface to their informational needs. Surprisingly, we show that on average an increase in information degrades trading performance. An explanation for this effect may lie in cognitive theory. Displaying more information increases the participants' cognitive load and hence may reduce decision accuracy and confidence. We are able to distinguish between trading behavior and performance and thereby provide insight into the interplay between interface, information and decision-making. Additionally, we can track the influence of individual information elements and identify those that improve or decrease trading performance. Focusing on the anchoring bias we link behavioral aspects of the market participants to the quality of their decisions. Creating a link between behavioral aspects of the participants and quality is important in that the quality of the predictive power is directly negatively affected if participants make systematically biased decisions.

In the second experiment participants can individually choose between two trading interface types. One interface type is a standard trading interface, whereas the other hides most market complexities. Recording through which interface an order is submitted allows us to link trading performance and interface type. Evaluating the hidden market design paradigm from an individual perspective, we find that alternative trading interfaces change participants' behavior. Using the trading wizards traders are more likely to submit market orders and submit orders with smaller sizes. Furthermore, and against naive intuition, we find that orders submitted through a simplified interface are more likely to be profitable compared to orders which are submitted through the default trading interface.
The remainder of this chapter is structured as follows: the second section presents a review of related work in the market interface domain. Additionally a short overview of the economic influence on decision behavior is given. The third section details the field experiment setting and the framing of the participants' trading process. The subsequent section first presents some descriptive data and then introduces the evaluation methodology. Specifically, we use market measures to separately analyze trading performance and trading behavior. In section five we link the interface types to trading outcome and interpret the results. Finally section six concludes this chapter.

7.2 Related Work

7.2.1 Market Interface Design

A fundamental assumption of many market designers is that participants are sophisticated and act rational. Hence, participants are able to express their expectations as bids and understand the underlying implications. As a consequence, designers have developed mechanisms that are theoretically efficient if participants are perfectly rational (Maskin, 2008). Assuming a perfect rational agent; designing the market interface is just a means of presenting the mechanism. In this line of reasoning oftentimes the market interface is used to present as much information as possible to support the trading process. However, it is well known that individuals are bounded rational and might not able to cognitively handle all available information (Simon, 1997). This raises the question of which information supports trading decision and which information distracts.

7.2.2 Hidden Market Design

Challenged by the rise of complex markets (e.g. Energy, P2P resource sharing) in which non-sophisticated users find it hard to interact, Seuken et al. proposed the idea of Hidden Market Design. *"The Hidden Market Design challenge is to find new techniques and approaches towards designing and building hidden markets for non-sophisticated users. The primary goal [..] is to find the right trade-off between hiding or reducing some of the market complexities while maximizing economic efficiency attained in equilibrium."* (Seuken et al., 2010). Hence, the goal is to lower the entrance barriers (e.g. market complexities) for non-sophisticated users to participate in markets. The simplification can be achieved by either changing the user interface or adapting the market rules. Following the idea they design a market-based P2P backup application (Seuken et al., 2010). In the paper they address both aspects, the user interface eliciting participants' preferences and the market rules, standardizing the market interaction. However, it remains unclear how the simplified trading interface affects market efficiency and individual trading decisions. The first step is a better understanding of which information elements support and which elements hinder the individual trading process. Moreover, as individuals have different informational needs and cognitive capacity, it seems fruitful to develop customized market interfaces.

7.2.3 User Behavior in Trading Environments

As the literature on interface design in financial markets is sparse, we mostly rely on findings from related fields. Kauffman and Diamond (1990) highlight the importance of research on behavioral decision making and information presentation effects. They examine how behavioral effects may become operative in screen-based securities and foreign exchange trading activities, where users

can choose among information presentation formats which support trader decision making. They present a model to identify where and how information, heuristics and biases might effect decision making in the trading environment. Besides this theoretical model there exists, to our knowledge, no work linking the decision making in continuous markets to the trading interface.
In the domains of decision support systems and online shopping environments the influence of the interface on decision behavior has been repeatedly demonstrated. Kleinmuntz and Schkade (1993) find that information displays influence decision processes by facilitating some decision strategies while hindering others. Decision makers balance the desire to maximize accuracy against the desire to minimize effort. Kleinmuntz and Schkade (1993) further separate characteristics of information displays into: (i) the form of individual items (numerical, verbal or pictorial), (ii) the organization into meaningful structures (groups, hierarchies or patterns) and (iii) the sequence (the order in which information element appears). In a follow-up study they show that organization strongly influences information acquisition while form influences information combination and evaluation. The sequence had only a limited effect on information acquisition (Schkade and Kleinmuntz, 1994). Investigating the relationship between problem representation and task type in information acquisition, Vessey (1991) develops the cognitive fit theory. The theory proposes that the correspondence between task and information presentation leads to superior task performance for individual users. In several studies, cognitive fit theory has provided an explanation for performance differences among users across different presentation formats such as tables, graphs, and schematic faces (Vessey, 1994; Vessey and Galletta, 1991). Additionally, they show that increasing interface flexibility instead of an informed choice of of display format may be harmful rather than helpful to the problem solver. Similarly Speier and Morris (2003) compare the use of visual and text-based interfaces for low and high complexity tasks. They find that in low complexity environments participants perform better using text-based query tools. However, in the high complexity environments participants perform better with visual support. Turning to the optimal pool of available information in decision support systems, empirical work has shown that users

can handle only a certain amount of data. Malhotra (1982) concludes that individuals cannot optimally handle more than ten information items or attributes simultaneously. Testing decision accuracy Streufert et al. (1967) show that as information load increases, decision making first increases, reaches an optimum (information load ten) and then decreases. Finally, in an interactive home shopping simulation Ariely (2000) tested how the participants' control over information influences their utilization of this information. He compared four settings; if information control was high-low and the task complexity was high-low. He finds that increased control over information leads to better performance in tasks with low complexity and lower performance in the high complexity setting. He reasoned that for participants in the low complexity setting, when demand on processing resources is low, more information is beneficial. In complex situations however the information is detrimental to performance due to the additional burden of selecting the right information (Ariely, 2000). He concludes that when cognitive load is high (when the task is novel or difficult) high information control can be harmful.
To summarize previous work, the amount and control of information, as well as the information representation does influence user behavior. On the one hand information control improves performance by improving the fit between actions and outcomes. On the other hand in terms of cost (disadvantages), information control requires the user to invest processing resources in managing the information amount and flow. As a conclusion, information control has both positive and negative effects on performance. The two tasks of processing and managing information are related and codependent. Finally, one must note that previous work has mainly investigated the topic in laboratory settings. We analyze decision behavior in a field experiment setting, namely a prediction market.

Anchoring

As in many domains of human judgment and decision-making, market participants rely on judgmental heuristics and mental shortcuts that turn complex decisions into simple judgment tasks. One of these heuristics or biases is the anchoring effect. It refers to the fact that a previously mentioned random number

influences the forecasting judgment (Camerer and Lowenstein, 2003). Kahneman and Tversky (1979) show that even when decision makers are anchored to an arbitrary number such as their social security number a following decision is influenced by this random number.
On one hand market interfaces provide public information to form a belief. On the other hand it offers the possibility to anchor the decision to some (misleading) value. Therefore, another question is whether the interface in general influences decision behavior. But although prices may have no memory, investors do. In fact, it has been demonstrated that past stock prices do influence forecasts of stock prices (De Bondt, 1993). Furthermore, Mussweiler and Schneller (2003) studied how charts depicting past stock prices influence investing decisions. They find that market participants buy more and sell less when the critical chart is characterized by a salient high rather than by a low. One finding from laboratory experiments is that individuals with low cognitive abilities tend to be significantly more affected by behavioral biases (Hoppe and Kusterer, 2010).

Influence on Market Efficiency

Psychologists have demonstrated a variety of systematic departures from "rational" decision making by individuals. These lead to substantial information processing or judgment biases and colored expectations (Camerer and Lowenstein, 2003; Oechssler et al., 2009). Despite the evidence for persisting biases, market prices have not been distorted (Gil and Levitt, 2007; Forsythe et al., 1999; Luckner, 2007). When the bias is publicly known markets provide an incentive to de-bias (Gruca and Berg, 2007). As long as there are enough rational traders actively compensating the bias, price accuracy is not affected. Cowgill et al. (2009) report that pricing biases declined over the sample period and their market performed better as collective trading experience increased. As in any setting where biased agents are involved, following Hahn and Tetlock (2005) the real question is whether markets are more robust to the participation of irrational agents than other mechanisms. So far there is no definite answer to this question.

7.3 Setting and Research Question

7.3.1 Experimental Setting

In order to test the effect of different (customizable) trading interfaces on trading behavior and performance we use data from the EIX prediction market for economic variables. The EIX provides three trading interfaces which are displayed in Figure 7.1.

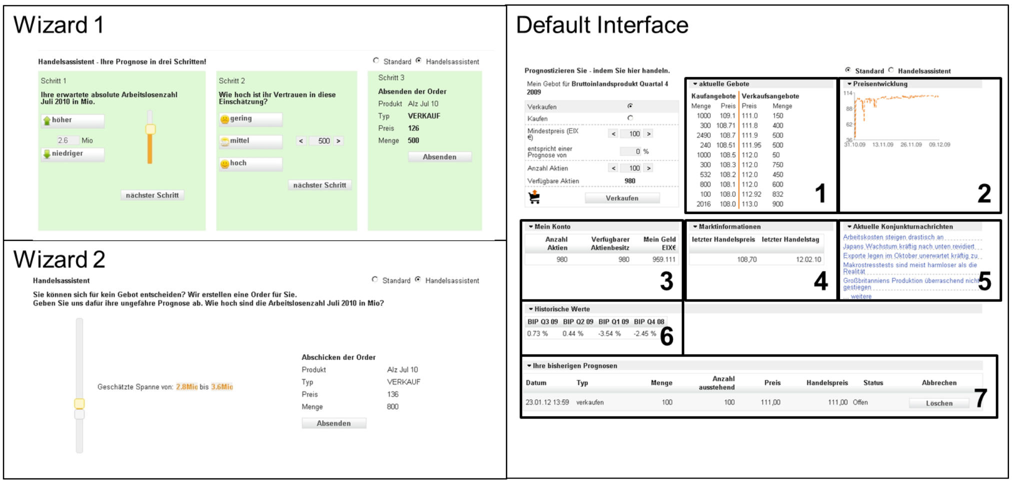

Figure 7.1: Three trading interfaces: Two wizards (left), default customizable trading interface (right)

Customizable Trading Interface

The default trading interface is displayed on the right side in Figure 7.1. Participants have convenient access to the order book (I1) with 10 levels of visible order book depth, the price chart (I2), the account information (I3) and market information (I4) such as the last trading day. As additional information the Handelsblatt provides access to an up-to-date economic news-stream (5) and finally the indicator's last years performance is displayed (I6). In the second round, we added a panel to display a list of previous orders (I7).

Participants are able to customize their trading interface individually. By clicking the small arrows the seven information panels open and close. In the default setting, only the trading mask and the seven headlines are visible. After each submitted order the chosen interface is saved per user. On user return the system opens the previously used interface elements on default. The advantage is twofold; first users have a convenient option to customize their trading experience, secondly we can assess which self-selected information pieces have influenced the participants' decision processes. Additionally we did not have to form groups with different interfaces and assign users to certain groups. This setting would possibly create an unfair experience.

Trading Wizards

Additionally to the default trading interface, participants have the choice to switch to a trading wizard guiding their trading decisions. In order to test for the interface influence on trading performance we designed two different wizards displayed in Figure 7.1 on the left hand side, marked with $Wizard_1$ and $Wizard_2$. Participants are randomly assigned in one of two groups with access to one of the two different trading wizards. Interface $Wizard_1$ is designed as a three step trading wizard, with three (green) boxes appearing in order. In the first step participants indicate if they believe the prediction to be higher or lower than the current market forecast. In the second step they are asked about their confidence in their prediction. The third box just displays the generated order. Interface $Wizard_2$ simply asks the participant to indicate a prediction interval with two handles. On the right hand side an order is automatically generated depending on the current orderbook and the distance between lowest and highest indicated prediction value. The interface is similar to and was inspired by the Yoopick interface (Goel et al., 2008). It is noteworthy that both wizards provide far less information than the default interface. In terms of Seuken et al. (2010) interface type $Wizard_1$ can be considered as a weakly hidden market interface, whereas type $Wizard_2$ hides the market completely .

7.3.2 Research Model

In order to answer the question on how to design (web-based) trading interfaces we have to deeply understand if and how different interfaces influence trading behavior and performance. More specifically, we need to analyze how participants search for information and how they incorporate this information in their trading process. Are they able to optimally utilized all available information? To give indications for these research questions we run two experi-

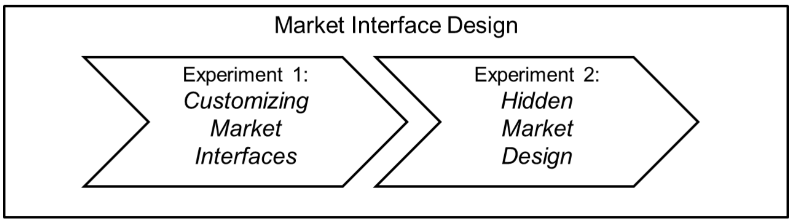

Figure 7.2: Market interfaces: Experiment overview

ments. As previously described, market participants can access the market via two trading venues. The default trading interface is fully customizable whereas the alternative trading venues (trading wizards) aim at supporting trading decisions. Following the experimental setup (see Figure 7.2), we start by analyzing how participants customize their interface and how this correlates with trading behavior and performance. In the second experiment we analyze how hiding certain market features affects participant performance. Hence, we evaluate the *Hidden Market Design* paradigm.

From an abstract perspective, both experiments and their evaluation follow the research model depicted in Figure 7.3. First traders decide to use a certain interface or customize their interface a certain way (H1). Correlated with this choice is their trading behavior. It seems likely that when using *hidden market* interface, certain market behavior is supported like submitting price taking orders. This will be analyzed in a second step (H2). Finally and most importantly by controlling for trading behavior we analyze how the resulting trading performance is influenced by different trading interfaces (H3). Generally as the research model depicts, we have to control for market micro-structure effects such as different contracts. The following two sections detail our research hypotheses for both experiment separately.

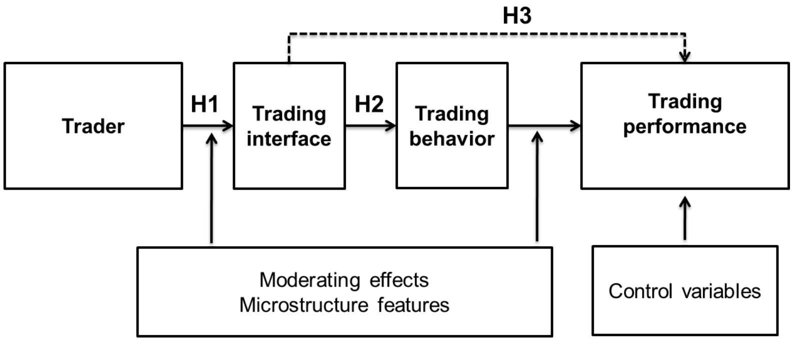

Figure 7.3: Research model

Customizing Market Interfaces

In the first step we analyze how participants individually customize their user interface. From another perspective we analyze which information different participants regard as useful. On one hand all information might help the user to trade better and improve her decisions. On the other hand no interface panel can be regarded as indispensable in order to trade. Therefore, we have first to analyze which interface elements are regarded worth considering in the trading process. Following Ariely (2000) we assume that participants choose different information elements as they try to adapt the interface to their informational needs. We expect users who are familiar with market environments to use more information elements. Users with no market experience might feel confused by too much data and hence reduce the interface to the simple basics.

E1-H1) *Users with a high market knowledge self-assessment use more information elements.*

In a second step (H2) we present how the self-chosen interface influences the participants' trading behavior. As all traders have the same start portfolio, the size of a trade is a proxy for the trader's confidence perception. Assuming that participants using more information (a high number of open information elements) are more confident about how to trade, it seems reasonable that the resulting order size is on average higher. Another individual market behavior is how participants submit their orders. We distinguish between market orders and limit orders. Market orders trade instantaneously against a standing limit order. Therefore, traders submitting market orders pay the effective spread in order to execute directly knowing that the order will be executed. Less confident traders want to keep the effective spread and submit limit orders. We assume that traders using more information are more confident about their decisions and submit market orders. Another question is whether the interface in general influences trading behavior. On one hand the interface provides public information to form a belief; on the other hand it offers the possibility to anchor the decision to some (misleading) value. Tversky and Kahneman (1974) show that even when decision makers are anchored to an arbitrary number such as their social security number a following decision is influenced by this random number. As a consequence we assume to find that when the historic value interface is open it induces an anchor effect. Accordingly, the hypotheses for the participants' trading behavior (H2) are:

> E1-H2a)*Participants with a high number of open information elements are more likely to submit orders with above average quantity.*
> *E1-H2b)The higher the number of open information interfaces the higher the chance that participants submit market orders.*
> *E1-H2c)When the historic value interface (I6) is open, the difference between the last historic value and the submitted limit price is lower.*

Finally and most importantly, we analyze how the self-chosen interface influences the participants' trading performance (H3). The intuitive assumption is the more information the better the decision accuracy. Hence a higher number of open information elements lead to a better trading performance. As presented, previous work suggests that decision performance might suffer if the information load is too high or the control of information distracts from the problem solution Malhotra (1982). Therefore, the alternative hypothesis is that too much information reduces decision accuracy.

> E1-H3a)*The more information elements are open, the better the participants' decision accuracy.*
> *E1-H3b)The more information elements are open, the lower the participants' decision accuracy.*

In combination, the three steps provide a first recognition of a market's interface impact on trader behavior. Moreover, they provide insight how a market's interface affect individual trading behavior and subsequently trading performance.

Hidden Market Design

We expect users who are familiar with market environments to use the default interface with more information. Users with no market experience might feel confused by too much data and might switch to a trading supporting interface. In the first step (H1) we present how the self-chosen interface influences the participants' trading behavior. As all traders have the same start portfolio the size of a trade is a proxy for the trader's confidence perception. Assuming that participants using the wizards are less confident about how to trade, it seems reasonable that the resulting order size is on average lower.
As in the previous section detailed another individual market behavior is how participants submit their orders. If they submit market orders they are confident about their forecast, if they submit limit orders they want to keep the realized spread. As the wizards do not display the current orderbook, it is reasonable to assume that wizard users are more likely to submit market orders. As a consequence the hypotheses for participants trading behavior (H1) are:

E2-H1a)*Orders which are submitted through a trading wizard are smaller in size on average.*

E2-H1b)*Using the trading wizard increases the chance that participants submit market orders.*

Finally and most importantly, we analyze how the self-chosen interface influences the participants' trading performance (H2). As more information is displayed in the default trading interface an intuitive assumption is to expect a better trading performance through the default interface. However an alternative perspective from decision theory is, that the more information, the worse the performance (Malhotra, 1982). Thus the hypotheses for the interface influence on trading performance (H2) are:

E2-H2a)*Using a trading wizard improves the participants' trading performance.*

E2-H2aa)*Using a trading wizard impairs the participants' trading performance.*

In combination the two steps provide a first empirical analysis of the hidden market design paradigm. Moreover, they provide insight how a market's interface effect individual trading behavior and subsequently trading performance.

7.4 Results

The following section first presents some descriptive market statistics and then details a regression framework measuring the effect of different trading interfaces on trading behavior and performance.

7.4.1 Descriptive Statistics

The following data includes the timespan from 30th October 2009 till 31st of October 2011. In total 1,235 participants registered at the EIX market, of those 824 submitted at least one order. We discard all stocks with less than 50 transactions. Altogether participants submitted 79,334 orders resulting in 34,028 executed transactions. Previous work showed that the market-generated forecasts performed well in comparison to the "Bloomberg"- survey forecasts, the industry standard (See chapter 4 and Teschner et al. (2011)).
For every order we record the open interfaces elements (I1-I7, see Figure 3.3). In the following an interface variable is 1 when the element is open otherwise it is 0. Moreover we record through which trading venue (wizard one, wizard two or default) the order was submitted. In the following an interface variable is 1 when the trading venue is used otherwise it is 0, e.g. variable W_1 is 1 if the alternative trading screen $Wizard_1$ is used (see Figure 3.3). In our field experiment we asked participants to self-assess their market-knowledge. According to their rating, we cluster the participants into two groups; the good (MK=1) and the not good (MK=0) market knowledge groups. In the first group are 499 participants and 763 participants rated their knowledge as not good.

7.4.2 Methodology

Analyzing Trading Behavior

As described in the last section to measure the participant's decision confidence we use two proxies. The order size and how traders submit their bids. In order to capture how predictor variables correlate with the submitted quantity we use OLS regressions.

$$\begin{aligned} Quantity_o = \quad & \alpha + \beta_{1..n} Predictor_{1..n,o} \\ & + \gamma_{1..m} Behavior\ Control_{1..m,o} \\ & + \delta_{1..i} Market\ Control_{1..i,o} \end{aligned} \tag{7.1}$$

We relate the quantity of a specific order to n predictor variables (e.g. to the number of open interfaces.) As the different indicators exhibit different historic variances, e.g. exports are much more volatile than inflation, we control by adding the market dummy variables MC_i. Similarly to control for the self-assessed market knowledge or other trader behavior we add m behavior control dummies. The control variables are included in all presented regressions.
For the second proxy we look at how users submit their offers. For an executed trade there are only two possibilities; either an order is a limit order or it is market order. The market order is initializing a trade against a standing limit order. As this is a binary outcome we use a binomial logistic regression. If a trade is initializing, which means it is market making, the dependent variable is 1 otherwise it is 0. Equation (7.2) exemplifies this. The regression measures the influence of predictor variables on the probability whether a trade is initializing or passive.

$$\begin{aligned} \log \frac{\pi_{Init_o}}{\pi_{Trade}} = \quad & \alpha + \beta_{1..n} Predictor_{1..n,o} \\ & + \gamma_{1..m} Behavior\ Control_{1..m,o} \\ & + \delta_{1..i} Market\ Control_{1..i,o} \end{aligned} \tag{7.2}$$

Measuring Trading Performance

As discussed in Chapter 2, we can ex-post measure the information content of each order. Moreover, the $Score_{o_i}$ (equation 2.8) rates an order as profitable or unprofitable. This allows us to capture the interface effect on an order per order basis.

For the profitability measures we use two regressions to capture the effects. First we measure the influence of predictor variables on the probability whether a trade is profitable or not.

$$\begin{aligned}\log \frac{\pi_{Score_o}}{\pi_{Trade}} = \quad & \alpha + \beta_{1..n} Predictor_{1..n,o} \\ & + \gamma_{1..m} Behavior\ Control_{1..m,o} \\ & + \delta_{1..i} Market\ Control_{1..i,o}\end{aligned} \tag{7.3}$$

The dependent variable is the score defined in equation 2.8 which is 1 for a profit and 0 for a loss. As before we control for different underlying uncertainties in the market categories and different trading behavior by adding dummy variables. Finally, we use OLS-regressions to estimate the effect on the resulting profit per order.

$$\begin{aligned}Profit_o = \quad & \alpha + \beta_{1..n} Predictor_{1..n,o} \\ & + \gamma_{1..m} Behavior\ Control_{1..m,o} \\ & + \delta_{1..i} Market\ Control_{1..i,o}\end{aligned} \tag{7.4}$$

As we basically use a panel data set (e.g., the EIX data set contains observations on multiple indicators from different individuals over time) OLS standard errors might be biased. Hence, besides controlling for indicators, we use participant clustered standard errors. More precisely, we are using Rogers (1994)-standard errors, which are White (1980, 1984)-standard errors adjusted to account for the possible correlation within a cluster. Hence, these are also called Rogers standard errors in the finance literature. In order to separate effects not only between participants but also within, we run the previously described regressions also as fixed effects models. This is identical to adding N-1 dummy variables, where N is the number of participants in the sample.

Following the experiment sequence, we first analyze how customizing trading interfaces affects trader behavior. We then compare trading between the three interfaces used.

7.4.3 Customizing the Trading Interface

We start by showing that participants choose different information elements to support their trading. Moreover, individual behavior differs depending on the interface elements implemented. Controlling for different trading behavior we find that a market participants using a lower number of open interfaces are more likely to submit profitable orders. Table 7.1 shows how market knowl-

	I1	I2	I3	I4	I5	I6	I7
Good	95%	38%	40%	48%	16%	43%	53%
Not Good	89%	33%	42%	44%	20%	40%	53%
Difference	6^a	5^a	-2^a	4^a	-4^a	3^a	0
(t-Stat.)	(30.8)	(9.9)	(-4.1)	(11.2)	(-16.7)	(6.9)	(0.8)

Table 7.1: **Market knowledge self-assessment and interface customization.** The table shows the interface usage (I1-I7) separately for the two user groups. (e.g. In 95% of all trades a trader with good market knowledge, the orderbook (I1) is open.) We test for the difference between the two groups with a simple t-test. The superscript 'a' denotes significance at the 0.1% level.

edge self-assessment and interface choice is related. Participants with a good self-assessed market knowledge use the orderbook (I1), the price chart (I2), the market (I4) and the historic values (I6) more often than the other participants. From another perspective, participants with a not good self-assessment open their account (I3) and the news (I5) more often. Turning to the first hypothesis we find that the number of open interfaces on average is higher (3.04 vs 2.89; $t - stat.: 8.9; p - value < 0.1\%$) for users with good market knowledge self-assessment. As a consequence we acept E1-H1.

7.4.4 Customized Interface and Trading Behavior

A common proxy for confidence in a trading environments is the submitted quantity. We suspected that participants with a high number of open information elements are more confident about their trading decision and thus submit orders with a higher quantity. As presented in Table 7.2, Model A; the opposite is the case; the higher the number of open interfaces the lower the submitted quantity. Accordingly we reject E1-H2a.

	Trading Behavior		**Trading Performance**	
	Quantity Estimate (t-Stat.) Model A	Order-type Estimate (χ-Square) Model B	Profit Estimate (t-Stat.) Model C	Score Estimate (χ-Square) Model D
Dependent variable	$Quantity_o$	$Init_o$	$Profit_o$	$Score_o$
Number of Interfaces	-84.26[a] (-12.2)	-0.06[a] (215.4)	1291.92 (0.8)	-0.04[a] (78.6)
Market Knowledge	-5.36 (-0.2)	-0.29[a] (213.4)	-590.06 (-0.1)	0.08[a] (17.9)

Table 7.2: **The influence of the number of open interfaces.** The Model A gives the values for the OLS Quantity regression (equation 7.1). The estimates show that if the number of interfaces is increased (e.g. from 2 to 3) the submitted quantity per order is reduced by 84. As Model B and D are Logit-regressions, the interpretation of the estimates is different. The estimates represent the change in the log odds of the outcome for a one unit increase in the predictor variable. (The chance that an order is a market order or that it is profitable is reduced with number of open interface elements.) The superscript 'a' denotes significance at the 0.1%, 'b' at the 1% level.

Investigating further which interface elements drive the quantity decision we use equation 7.2, regressing the submitted quantity on individual information elements. Table 7.3, Model A, shows the results. The orderbook, the price chart, account information and the previous order elements have an increasing effect.

We assumed that participants with more information would be more confident and hence submit more market orders. As the estimates in in Table 7.2, Model B; show this is not the case. The higher the number of open information interfaces the lower the chance that participants submit market orders. We reject E1-H2b and conclude that a high number of open interfaces results in a higher chance that participants submit limit orders. Following our arguement, one can interpret that participants with a higher number of interfaces act more cautious and submit limit orders, in order to keep the possible realized spread. Looking at how individual interfaces influence the order-type; Table 7.3, Model B, shows

that only an open news interface (I5) correlates with a higher chance of submitting market orders. The market information element (I4) renders the participants more cautious.
Taking the two proxies together, we conclude that trading confidence is reduced the more information the participants use. Alternatively one might argue that uncertain participants, with low confidence, use more information elements.
Our last behavioral hypothesis states that the historic value interface (I6) induces an anchoring bias, as participants orient their trading decision on an easy to understand and accessible variable. This simple forecast heuristic does not imply that the decision is necessarily bad. As Osterloh (2008) shows, the naive forecast[1] is often as good as the expert prediction for economic indicators.
In order to analyze whether participants exhibit an anchor bias we use the following OLS regression.

$$\Delta_{o,i} = \alpha + \beta I_{6,o} + \gamma MK + \sum_{j=1}^{5} \delta_j M_j \tag{7.5}$$

$$with\ \Delta_{o,i} = |Limitprice_{i,o} - Historic\ Value_i| \tag{7.6}$$

On the left hand side is the distance between submitted limit price and the last historic value which is displayed in the interface element I6. Additionally we control for the market knowledge and the risk factors from the different markets. If participants exhibit an anchor bias the last historic value is used as an orientation and hence the distance $\Delta_{o,i}$ should be reduced. Table 7.4 depicts that the distance between historic value and limit price is significantly reduced if the corresponding interface (I6) is open. We follow that the display induces an anchor effect and accept E1-H2c. Hence, taking these three results together, we find that the trading interface at least partly influences how participants submit their orders.

7.4.5 Customized Interface and Trading Performance

We suggested two alternative hypothesis regarding the interface influence on decision accuracy. One might intuitively suspect that a higher number of information panels correlates with better trading decisions. Turning to Table 7.2, Model C; reveals that the chance of submitting a profitable order is lower with an increasing number of information panels. We thus reject hypothesis E1-H3a and accept E1-H3aa. Interestingly looking at the interface elements supporting successful trading, it turns out that the price chart and the previous order element have a positive effect (Table 7.5). Moreover, the market element (I4) has a negative effect. A possible explanation for this result might be that certain information provided by the system may not actually help in the decision making process. When designing the interface the goal was to support the participants to make good forecasts and consequently make good decisions. As the really informative interface elements can only be identified ex-post this result would suggest to rework the interface design and reduce the number of information elements to a minimum.

7.4.6 Interface Type and Trading Performance

Turning to the second experiment we will evaluate in this section how two alternative interfaces support non-sophisticated traders participating in a (complex) prediction market. We show how individual behavior differs depending on the interface used. Controlling for trading behavior, we find that market participants using a trading wizard are more likely to submit profitable orders. Following the presented research model we start by analyzing how trader behavior differs if participants use certain interfaces.

A common proxy for confidence in a trading environment is the submitted quantity. We assumed that participants using the wizards are less confident about trading, and hence the resulting order size is lower on average. We test these

[1] using the previous value as a predictor

hypothesis using following OLS regressions.

$$Quantity_o = \alpha + \beta_1 W_1 + \beta_2 W_2 + \gamma Init. + \delta TD. + \lambda MK. + \sum_{i=1}^{5} \phi_i M_i \qquad (7.7)$$

In Table 7.6 the results for regression 7.7 (Model A and B) are depicted. Model B includes participant fixed effects. Hence, the estimate show the effect within individual participants. Participants using the wizards submit orders with a lower quantity of at least -569 per order on average. Thus, we can accept hypothesis E2-H1a. Separating the effect for certain wizard types, we see that the results are stronger for wizard type W_2.

$$\log \frac{\pi_{Init}}{\pi_{Trade}} = \beta Wiz. + \gamma Init. + \delta TD. + \lambda MK. + \sum_{i=1}^{5} \phi_i M_i \qquad (7.8)$$

For the second proxy we look at how users submit their offers. Orders are either liquidity taking or liquidity providing. We code liquidity taking (initializing) orders with $Init. = 1$. Equation (7.8) measures the influence of the interfaces on the probability whether a trade is initializing or passive. As this is a binary outcome we use a binomial logistic regression. We assumed that participants using the wizard do not see the orderbook and hence are more likely to submit market orders. As the estimates in in Table 7.6, Model C and D; show this is the case. Accordingly we accept E2-H1b. Again looking at the particular influence of each interface we find that the results are stronger for the wizard type W_2. As a conclusion, we find that participant behavior changes when using alternative trading interfaces.

We suggested two alternative hypotheses regarding the interface influence on trading performance. One might intuitively suspect that more information on the default trading interface leads to better trading decisions. In that sense a naive assumption is that the participants using the default trading interface are more likely to submit profitable orders. To measure the effect on trading performance we adapt equations (7.8) the following way; we exchange the dependent variable $\log \frac{\pi_{Init}}{\pi_{Trade}}$ by $\log \frac{\pi_{Score}}{\pi_{Trade}}$. The *Score* of an order rates if the order moves the price in right direction. However, we also want to know if the profit per order is increased. Hence we additionally run OLS regressions on the profit per order.

$$Profit_o = \alpha + \beta_1 W_1 + \beta_2 W_2 + \gamma Init. + \delta TD. + \lambda MK. + \sum_{i=1}^{5} \phi_i M_i \quad (7.9)$$

The results are depicted in Table 7.7. As Model C and D are Logit-regressions, the estimates represent the change in the log odds of the outcome for a one unit increase in the predictor variable. In the random effects model (C) we find no difference between using either one of the wizard or the default trading interface. However, turning to the fixed effects model (D), we find a -within subject-effect. Participants using the second wizard are more likely to submit profitable orders. For the first wizard we find no difference compared to the standard interface. The Models A and B in Table 7.7 present the estimates for equation 7.9. The results show that participants using the second wizard gain a higher profit (5,334 currency units). Participants using the first wizard loose money compared to using the standard market interface. In the fixed effects model (B) the direction still holds. However, the estimates are not significant.

Taking the results together, we conclude that alternative trading interfaces can help improving trading performance. We thus reject hypothesis E2-H2aa and accept E2-H2a. A possible explanation for this result might be that certain information provided by the system may not actually help but impair the trading decision process. It seems that the results are stronger for interface W_2 the strongly hidden market interface. We argue that one has to be careful designing such interfaces. However, as the interfaces are merely a first attempt to design simplified interfaces, there is a reason to believe that there is plenty of room for future improvement.

7.5 Conclusion

Various allocation problems call for market based solutions. However, market complexities impose high entry barriers for non-sophisticated users. One reason is that in markets preferences are usually communicated through bids and offers which requires participants to adapt to a different mental model. Recently researchers proposed the idea of hidden market design which merges the fields of market design with user interface design in order to make complex markets accessible to a broader audience. As more trading decisions are facilitated through (web-based) "trading support systems" one of the most urging questions is how to design such interfaces. Moreover, it is important to design such interfaces without reducing market efficiency and individual trading performance.

7.5.1 Market Interface Design

In our field experiment participants trade in a complex prediction market which closely reassembles trading in financial markets. As the outcome of events in prediction markets is finally known, we can ex-post measure the participants' trading performance. In our market setting participants can individually customize their interface. We show that participants choose different information elements depending on their self-rated market knowledge. Apparently market participants try to fit the interface to their individual problem representation. Yet the individual motivation of the interface choice remains unclear.

We show that more information does not improve decision making but rather leads to decreasing trading performance. From the results presented, one might follow that information accessible on interfaces does not help forecasting economic variables and hence the participants' decision-making. Another interpretation, which is in line with previous work, may be found in cognitive theory. Too much information increases the participants' cognitive load and hence may reduce decision accuracy and confidence. Thus a high number of interface elements increases complexity and distracts from good decision making. As participants are able to customize their interface they seem unaware of the negative influence of the interface on their decisions. As a consequence market designers should not only limit the amount of presented information but also make a validated guess about which information is useful. Due to the close relation to decision processes this paper helps to understanding the impact of decision support system interfaces on decision-making in general. From previous work it is known that the amount and control of information, as well as the information representation does influence user behaviour. On one hand information control improves performance by improving the fit between actions and outcomes. On the other hand information control requires the user to invest processing resources in managing the information amount and flow. Therefore, information control has both positive and negative effects on performance. Previous work has mainly investigated the topic in laboratory settings. We analyzed decision behaviour in a field experiment setting. This work also provides insight into the interplay between interface, information and trading behavior. Furthermore, we hope this work is a good starting point for practitioners and researchers designing markets and their interfaces.

7.5.2 Hidden Market Design

Despite the rise of complex markets, non-sophisticated users still find it hard to interact with such markets. Evaluating the hidden market paradigm from an individual perspective, we find that alternative trading interfaces change participants' behavior. Using the trading wizards, traders are more likely to submit market orders and submit orders with smaller sizes. Against naive intuition, we find that orders submitted through the strongly hidden market interface are more likely to be profitable compared to orders submitted trough the default trading interface. A reason for that may be found in cognitive theory. Market complexity increases the participants' cognitive load and hence may reduce trading performance and confidence. As a result this work provides insight into the interplay between market design, interface, and trading behavior. Specifically in the domain of financial markets it is the first work to show the influence of the trading interface on trading behavior and performance.

	Trading Behavior	
	Quantity	Order-type
	Estimate	Estimate
	(t-Stat.)	(χ-Square)
	Model A	Model B
Dependent variable	$Quantity_o$	$Init_o$
I1	377.34^a	-0.04
	(5.1)	(0.59)
I2	297.52^a	0.04
	(8.2)	(1.17)
I3	167.9^b	-0.05
	(3.5)	(0.75)
I4	-403^a	-0.25^a
	(-8.4)	(21.15)
I5	-351.62^a	0.3^a
	(-7.5)	(36.22)
I6	-319.07^a	0.04^b
	(-5.59)	(5.83)
I7	204.88^a	0.01
	(5.0)	(0.01)
MK	-68.57^c	-0.27^a
	(-2)	(58)

Table 7.3: **The effect on trading behavior of customizing interfaces.** Model A gives the values for the interface-quantity regression. The estimates show that if a specific interfaces is open; how the submitted quantity per order is affected (e.g. If the orderbook (I2) is open the submitted quantity is increased by 297). As Model B is a Logit-regression the interpretation of the estimates is different. The estimates represent the change in the log odds of the outcome for a one unit increase in the predictor variable. (The chance that an order is a market order and that is profitable is reduced/increased with a specific interfaces being open.) The superscript 'a' denotes significance at the 0.1%, 'b' at the 1% level.

Dependent variable	Estimate (t-Stat.) $\Delta_{o,i}$	
I6	-2.41[a]	(-36.8)
Market Knowledge	-0.62[a]	(-7.8)

Table 7.4: **Anchor heuristic.** The tables presents the results from equation 7.6. The estimates show that if interface I6 is open, the difference between the submitted price and the last historic value is reduced on average by 2.4 price points. As the average price of stocks is around 120, the 2.4 make an economic impact (2 %). The superscript 'a' denotes significance at the 0.1% level.

	Profitability			
	Estimate (t-Stat.) Model A	Estimate (t-Stat.) Model B fixed effects	Estimate (χ-Square) Model C	Estimate (χ-Square) Model D fixed effects
Dependent variable	$Profit_o$	$Profit_o$	$Score_o$	$Score_o$
I1	-25299 (-1.4)	3360 (0.71)	0.20^a (7.42)	-0.38^b (5.26)
I2	14829 (1.25)	14628^a (3.78)	-0.37 (93.54)	-0.37^a (56.51)
I3	19921^b (2.14)	19558^a (3.8)	-0.08 (1.97)	0.15^b (4.21)
I4	-9808 (-1.21)	-10415^c (-2.03)	-0.11^c (4.05)	-0.03 (0.20)
I5	-8001 (-0.48)	-8208 (-1.63)	0.01 (0.08)	0.14^c (5.25)
I6	-3110 (-0.6)	-1540 (-0.39)	0.04 (0.99)	-0.01 (0.04)
I7	10704 (1.25)	10793^b (2.48)	0.11^b (5.56)	-0.04 (0.38)
MK	4683 (0.47)	- (-)	0.20^a (32.85)	- (-)

Table 7.5: **The effect of customizing interfaces on performance.** For Model A and B the estimates show that if an specific interface is open; how the resulting profit (in EIX currency units) is affected. As Model C and D are Logit-regressions the interpretation of the estimates is different. The estimates represent the change in the log odds of the outcome for a one unit increase in the predictor variable. The chance that an order is profitable is reduced/increased with a specific interfaces being open. The superscript 'a' denotes significance at the 0.1%, 'b' at the 1%, and 'c' at the 5% level.

	Trading behavior			
	Estimate (t-Stat.) Model A	Estimate (t-Stat.) Model B fixed effects	Estimate (χ-Square) Model C	Estimate (χ-Square) Model D fixed effects
Dependent variable	$Quantity_o$	$Quantity_o$	$Init_o$	$Init_o$
W1	-569^a (-6.3)	-514^a (-3.8)	0.15^b (3.2)	0.17^a (7.1)
W2	-761^a (-7)	-763^a (-4.5)	0.22^a (6.2)	0.22^a (7.2)
Init.	-152 (-1.6)	-82^b (-3.2)	- (-)	- (-)
TD.	130 (1.3)	198^a (7.7)	0.02 (1.1)	0.04^a (9.2)
Quantity	- (-)	- (-)	0 (0)	-0.01^b (3.2)
MK.	-40 (-0.2)	- (-)	-0.07 (1.5)	- (-)

Table 7.6: **Influence of trading wizards on trading behavior.** Model A gives the values for the regression (7.7). The estimates show that if a trading wizard is used, the submitted quantity per order is reduced by 569 for wizard one and 761 for wizard two. The results hold when using a fixed effects model (Model B).
As Model C and D are Logit-regressions, the interpretation of the estimates is different. The estimates represent the change in the log odds of the outcome if the predictor variable is one. (The chance that an order is a market order is increased when participants use either wizard.). Again the results hold when using a fixed effects model (Model D). The superscript 'a' denotes significance at the 0.1%, 'b' at the 1% , 'c' at the 1% level.

	Profitability			
	Estimate (t-Stat.) Model A	Estimate (t-Stat.) Model B fixed effects	Estimate (χ-Square) Model C	Estimate (χ-Square) Model D fixed effects
Dependent variable	$Profit_o$	$Profit_o$	$Score_o$	$Score_o$
W1	-3161.68^a (-3.3)	-2894.11 (-0.6)	-0.04 (0.2)	0.02 (0.1)
W2	5334.65^a (4.2)	5355.27 (0.9)	0.14 (1.2)	0.57^a (8.9)
Init.	-859.66 (-1.0)	-540 (-0.6)	-0.17^a (79.9)	-0.12^a (32.5)
TD.	-14278.10^a (-7.7)	-13978.68^a (-16.6)	-0.70^a (1441.2)	-0.62^a (920.4)
Quantity	-0.73 (-0.7)	-0.69^a (-4.5)	-0.01^c (3.5)	-0.01^a (10.0)
MK.	1511.67 (1.5)	- (-)	0.17^a (69.6)	- (-)

Table 7.7: **Influence of trading wizards on order profitability.** As Model A and B are Logit-regressions, the estimates represent the change in the log odds of the outcome for a one unit increase in the predictor variable. (The chance that an order is profitable is higher when participants use the second wizard.) The Models B and C present the estimates for the OLS Profit regression (equation 7.9). The estimates show that participants using the second wizard gain a higher profit (5.334). Participants using the first wizard loose money compared to using the standard market interface. In the fixed effects model (D) the direction still holds. However the estimates are not significant. The superscript 'a' denotes significance at the 0.1%, 'b' at the 1% , 'c' at the 1% level.

Chapter 8

Conclusion & Outlook

8.1 Contributions

MACROECONOMIC forecasts are used extensively in industry and government even though the historical accuracy and reliability is disputed (Osterloh, 2008; Schuh, 2001; McNees, 1992). We developed and study a prediction market specifically designed to forecast economic indicators such as GDP, Ifo-index, inflation, investments, export and unemployment figures in Germany. Analyzing the market's forecast accuracy, we find that forecast performance improves constantly over time and that generated forecasts perform well in comparison to the Bloomberg-survey forecasts, the industry standard.

The main goal of this work is to address the following five research questions:

- How to quantify prediction market performance?
- How to design a market to forecast macroeconomic indices?
- What is the effect of short-selling on market quality and forecast performance?
- How do incentives and feedback mechanisms affect participation in prediction markets?
- How do trading interfaces affect trading behavior and performance?

Given the results of this work, these can be briefly answered as follows.

Contribution 1: How to quantify prediction market performance? In the last 20 years over 150 articles on prediction markets were published (Tziralis and

Tatsiopoulos, 2007). However no meta-analysis was conducted that analyzed prediction markets accuracy for various types of problems and fields of application. One reason is the lack of a unified, stringent methodology used to analyze prediction market data. Although there are a some published guidelines (e.g., Plott and Chen, 2002; Luckner, 2008; Sripawatakul and Sutivong, 2010) on how to design and implement a prediction market, no work has yet described and summarized evaluation methodologies. Moreover, there is no general understanding what defines prediction market quality. Up till now, prediction markets are often quantified solely through the correlation between predictions (derived from prices) and event outcomes. Simple correlations ignore external effects such as uncertainty underlying the event outcomes. Hence a prediction market evaluation based only on forecast precision hinders the understanding of the key success factors.

We provide a methodology framework that places several prominent measures of market quality into the context of internal design factors. Based on a detailed literature review, the framework summarizes and applies finance, forecasting and prediction market literature. The proposed framework for prediction market quality consists of the four sections: liquidity, information, activity, and forecast performance. The first two sections describe finance methods such as spread estimates, price impact and probability of informed trading. The last two sections are based on forecasting and prediction market literature with a strong focus on relative forecast performance measurement. Using the EIX-market dataset, we apply and evaluate all described methods.

This unified methodology framework enable us to address questions regarding the effect of different market designs, the question if an increasing number of traders is always beneficial and how the application domain affects forecast accuracy. Moreover it provides the methodical foundation for the following research questions.

Contribution 2: How to design a market to forecast macroeconomic indices?

We first summarized findings from previous markets in the domain of macroeconomic forecasting and detailed the known shortcomings of the currently used

binary payout designs. We proposed a radically different approach using a linear payout function. The theoretical improvements are threefold: (i) the number of traded stocks is reduced leading to higher liquidity in the traded stocks, (iii) the "partition-dependence" bias can be avoided and (iii) information can be aggregated continuously and over longer time horizons.
We then designed and implemented a prediction market to forecast macroeconomic variables in Germany. The market acts as a mechanism not only to aggregate dispersed information but also to aggregate individual forecasts. It does so by incentivizing participation and rewards early, precise forecasts. Moreover, the market-platform is yet alone in aggregating these forecasts continuously and for a long time horizon. Turning to the market-generated forecasts, we find that forecast accuracy improves constantly over time as the forecast error drops by 0.012 points per day. Comparing the market forecasts to the Bloomberg-survey forecasts, we find that they perform equally well. However, the Bloomberg forecast is only available 8 days before the data release whereas the the market provides as accurate forecasts 30 days in advance. Additionally we are able to show that the market has three supplementary benefits. To begin with, we can show that the market measures can be used to predict the forecast error. This might enable us to enhance the market forecast by providing a measure for forecast confidence. Secondly, market measures can be used to identify valuable user input and forecast experts in near real-time. Detecting such input might possibly enable us to improve the information aggregation mechanism and the forecast performance of such systems. Finally, in line with previous work, forecasts can be improved by 16.7 % when combining Bloomberg-survey and market generated forecasts.

Contribution 3: What is the effect of short-selling on market quality and forecast performance? We first discuss how short selling can be implemented in the EIX market. Using an event study approach we then document the positive effect of short selling on market accuracy. After the introduction of short selling, forecast errors drop by -30.8 %. Finally, we find that spreads which are a proxy for market uncertainty are lowered by -51.3 %. The results indicate that short selling in linear prediction markets is an essential characteristic to enable high prediction accuracy. It remains unclear whether different short selling implementations lead to different results.

Contribution 4: How do incentives and feedback mechanisms affect participation in prediction markets? We presented an incentive scheme well-suited to motivate participants contributing their information for longer time horizons. Investigating the level of participation, we find that activity is mainly driven by a weekly newsletter which acts as a reminder. On days the newsletter is sent out we measure an 62 % increased trading activity. Assuming that classical feedback mechanisms would lead to different participation levels, we find that the induced competitiveness of market environments seem to superpose classical feedback mechanisms.
Our semi-anonymous game enables naïve and professional forecasters to test their forecast ability and compare it to their peers. We show that with every order a trader submits, her average profit per order increases by 1.6 currency units. This means that participants gain experience over time which indicates that the active engaging environment fosters learning. Testing if participants are able to learn their forecasting ability, we display a specifically designed forecast performance ranking. We show that participants who are low performing receive the necessary feedback and hence realize their low forecast ability.

Contribution 5: How do trading interfaces affect trading behavior and performance? In the EIX market setting, participants can customize their interface individually. We show that market participants try to fit the interface to their individual problem representation. Yet the individual motivation of the interface choice remains unclear. We can further show that more information does not improve decision making but rather leads to decreasing trading performance. It seems that too much information increases the participants' cognitive load and hence reduces decision accuracy and confidence. As participants are able to customize their interface they seem unaware of the negative influence of the interface on their decisions. As a consequence market designers should not only limit the amount of presented information but also make a validated guess about which information is useful.

Evaluating the hidden market paradigm from an individual perspective, we find that alternative trading interfaces change participants' behavior. Using the trading wizards, traders are more likely to submit market orders and submit orders with smaller sizes. Against naïve intuition, we find that orders submitted through the strongly hidden market interface are more likely to be profitable compared to orders submitted trough the default trading interface. Even though the wizard interface is mainly used by inexperienced participants the results show that participants using the wizard gain a higher profit of 5,334 currency units per order. As a result this work provides insights into the interplay between market design, interface, and trading behavior.

8.2 Complementary Research & Future Work

The following section presents some on-going and future work which mainly builds upon the thesis results.

A prediction market meta-study The studies available to date show high accuracy of prediction markets for various fields of application. However, some studies only report absolute accuracy or compare the results to benchmarks like polls or betting odds. To gain a better understanding of what drives market accuracy as well as market quality it seems promising to combine data from various field experiments. Since the emergence of the field, no meta-analysis was conducted that analyzed prediction markets accuracy for various types of problems and fields of application. This is also partly due to a lack of a common methodology basis. Hence, the presented prediction market quality framework enables us to use a unified methodology. A meta-study could tackle questions regarding the effect of different market designs, the question if an increasing number of traders is always beneficial and how the application domain affects forecast accuracy. Related to this question is, how traders' personality traits influence their decision making and if pre-selecting certain traders could improve market accuracy.

Analyzing trader behavior and personality traits Psychologists have demonstrated a variety of systematic departures from *rational* decision making by individuals. These lead to substantial information processing or judgment biases and colored expectations (Forsythe et al., 1999). Markets suffer from biases as well and it is an ongoing debate to which extent they affect market efficiency (Arrow et al., 2008). Objectively irrelevant (Huber et al., 2008) and selectively presented information (Dittrich et al., 2005) influence individual trading behavior. A promising approach to describe and explain financial decision making may be the explicit consideration of psychological factors. In particular heuristics and biases need to be integrated. Future work should link behavioral aspects of the market participants and the quality of their decisions. Creating a link between behavioral aspects of the participants and quality is important in that the quality of the predictive power might be directly negatively affected if participants make systematically biased decisions. This is relatively well known, but still not well understood or studied, hypothesis of the behavioral finance literature. A prediction market setup is well-suited to study the behavioral aspects of decision making because, in contrast to financial markets (i) the value of shares in our market is ultimately known and (ii) we can measure the participants' ex-post trading performance.
As discussed in Chapter 7 especially inexperienced traders gain from alternative trading interfaces. Moreover they perform well if they do not need to trade but rather submit their forecast expectation directly. Hence it seems fruitful to apply findings from the survey-design domain in designing such interfaces.

Confidence polling - extracting probability distributions from non-experts People have massive amounts of information about upcoming events, including: economic, financial, and political events. However, they are not all participating in the EIX market due to lack of understanding and market complexities. Building on findings in the last chapter, the question arises how to build a more efficient way for users to interact with the market. Most non-experts are familiar with polling-based interfaces. However, most polling interfaces only allow participants to submit point estimates. This approach however ignores current literature on overconfidence and information retrieval.

We could build a new web-interface that might be a more efficient method of gathering individual-level information than the currently utilized methods. The web-interface could capture the users' confidence range at self-selected intervals. They first set a range for the answer then provide a confidence level for that range. We then could utilize the mid-point of the confidence range as a point estimate. Assuming normality we could create a full probability distribution for each user. The confidence of the users, the inverse of the variance, might provide us with more meaningful information than just their point-estimate. Figure

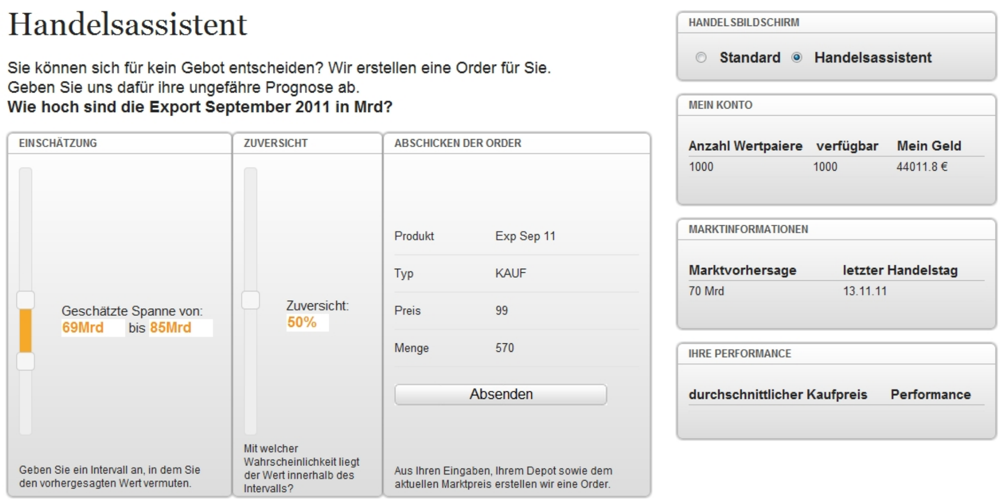

Figure 8.1: Polling market interface: Extracting full probability distributions from non-experts

8.1 displays such a setup. The interface could be designed as a three step (trading) wizard, with three boxes appearing in order. In the first step, participants indicate a forecast range to the given economic indicator. The default value is set to the current market forecast. In the second step they are asked to state the probability that the outcome is within the specified range. The third box just displays the generated order. The panels on the right hand side provide the participants with additional information, such as the data release date, the current market forecast and the participant's portfolio. Based on the results in Chapter

7, the wizard provides far less information than the default trading interface. The setup is an extension on the previous chapter on trading interfaces. The participants' full distributions can be used to submit orders in an efficient way. Aggregated confidence-weighted forecasts might be created several ways with this data. EIX field experiment is a perfect test environment for such this new interface. Another part of the EIX field experiment is a mobile application.

Mobile market interfaces With the enormous success of the iPhone platform a rising number of mobile applications is now available. Many of these applications allow users to make business relevant decisions on the fly. The question arises if and how decision behavior and decision performance is different in mobile market environments compared to stationary settings. In order to design mobile systems which support good decision making we need to analyze how participants search for information and how they incorporate this information in their decision process. Moreover we need to link behavioral aspects of users and the quality of their decisions in order to improve the design of mobile interfaces. A part of the EIX-platform is a mobile application for the iPhone. It seems worth pursuing to run field experiments to answer these questions. From a market engineering perspective, this research might also yield insights into the question of how mobile trading interfaces need to be designed. If participants submit expectations, the market designer needs to develop heuristics to translate the expectations into orders. One feasible way is to use the participants' expectations as input for automated trading agents.

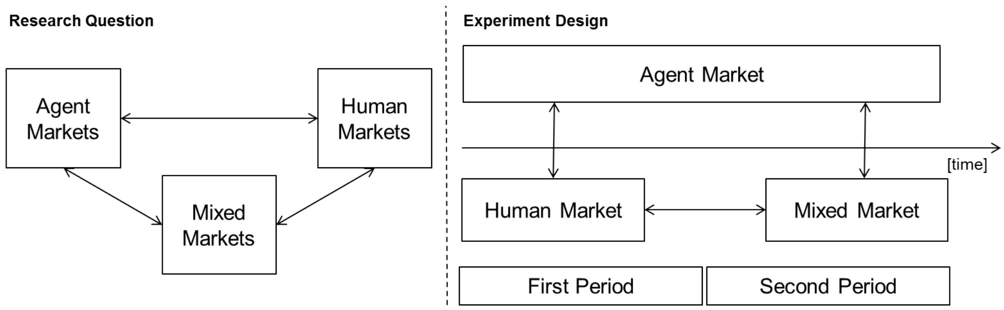

Figure 8.2: Algorithmic trading and market efficiency

Combining human and algorithmic agents for making predictions A recent experimental study by Nagar and Malone (2011) shows that adding automatic trading agents to a human prediction market might improve forecast performance. Measuring both market quality and forecast performance we could directly test if adding trading agents enhances market efficiency. This could also potentially enrich the current debate about limiting algorithmic trading in real financial markets. In general, we would like to compare three different parallel markets (Figure 8.2, left); a human only market, an agent only market and a mixed market in which both human and algorithmic trader interact. Hence, we have three comparisons to make. Using the EIX-market as a testbed we can design a field experiment to make these three comparisons. As depicted in Figure 8.2 (right side); we can compare the EIX to an agent market by running them in parallel (first period). By allowing agents to access the EIX at a specific time point we then can compare the EIX human market to an EIX hybrid market using an event study approach. This also gives us the third comparison between a hybrid market and an agent only market.
The promising EIX forecast results give rise to the idea to use prediction markets not only to aggregate information but to directly base decision upon market prices.

Prediction markets for making decisions To make an informed decision a decision maker must understand the likely consequences of their actions. In order to gain this understanding, in a typical scenario a decision maker consults a number of experts. To elicit experts' private information, the decision maker might run conditional decision markets, i.e. prediction markets with two distinct features: Firstly, the prediction is not an end in itself but there exists a direct link from predictions to a decision. Secondly, the market payoff is conditional on the decision maker's action. Only if an action is taken the true future state of the world can be observed the corresponding market is paid out. The other markets are voided. Experts hold private information, are self interested, and might have two (partially conflicting) incentives: On the one hand they can capitalize on their private information by trading in the decision markets; on the other hand they might have preferences over the principal's action. It seems fruitful to empirically study these decision markets in online or laboratory experiments.

Combining prediction markets and delphi-systems The track record of prediction markets suggests that markets may help to better foresee future developments and trends. However, looking at the range of applications, it becomes clear that there are certain limits. Complex forecasts, such as conditional or qualitative judgments are better gathered with traditional forecast methods such as delphi-systems. However, traditional delphi-systems have some known drawbacks. First of all, the success of a delphi-study largely depends on the participant selection (Ammon, 2009; Gordon, 2007). The most common selection criteria is reputation which is based on perceived expertise. However, Tetlock (2005) shows that perceived expertise does not correlate with individual forecast accuracy. The second drawback is the decreasing participant motivation over the study's course. The long, rigid and tedious process leads to decreasing participant numbers (Cuhls, 2003).
By combining prediction markets with delphi-studies there might be potential to reduce these drawbacks.

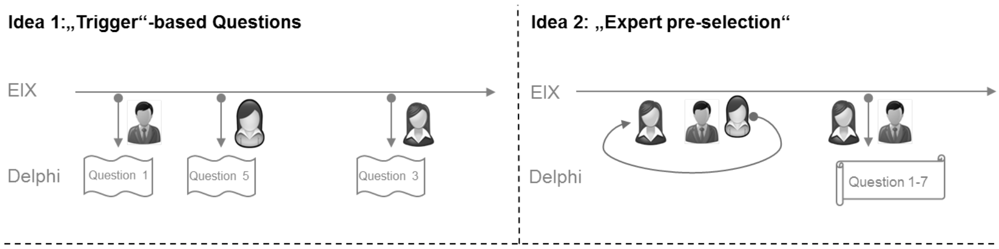

Figure 8.3: Combining prediction markets and delphi-studies

There are at least two ways delphi-studies can benefit from an accompanying prediction market; motivation and pre-selection of experts. Prediction markets motivate participants to contribute continuously through incentives and by providing constant feedback, both on the aggregate and the individual level.
More participants might be willing to participate (at least partly) in a delphi-study if they have indicated that they have information regarding a the topic. The question is how to figure out when a participant has information about a

topic. This can be detected through the prediction market. If participants change the market price, they have information about a certain topic and might be willing to fill out some related qualitative and possibly more complex questions (Figure 8.3, left side). As shown in the third chapter, individual forecast input can be measured and objectively evaluated in prediction markets. Hence, this might help to pre-select experts not based on their reputation but on their previous forecast performance (Figure 8.3, right side).

8.3 Summary

In this chapter we first presented the five main research contributions. The proposed unified prediction market framework provides the key elements to systematically analyze prediction market data. Using the EIX market data we showed the predictive power of prediction markets in the domain of macroeconomic forecasting. In summary, the EIX provided us with an ideal test-bed to (i) evaluate a novel linear market design, (ii) test the effect of market microstructure changes on forecast accuracy and (iii) and analyze the interplay between market interface and trading behavior.

Finally, we outlined seven main challenges for future research on prediction markets and market interfaces. In particular, we identified the combination of different forecasting systems, the incorporation of personality traits and the design of market interfaces. While the interface has gained only little attention in market engineering literature so far, the results of this thesis indicate that it plays a significant role in engineering efficient markets. We hope our approach will positively impact the (prediction) market design community and forecast results will eventually support economic policy-making in Germany by providing continuous information about the state of the economy.

Appendix A

Appendix

Figure A.1: Revision Screen

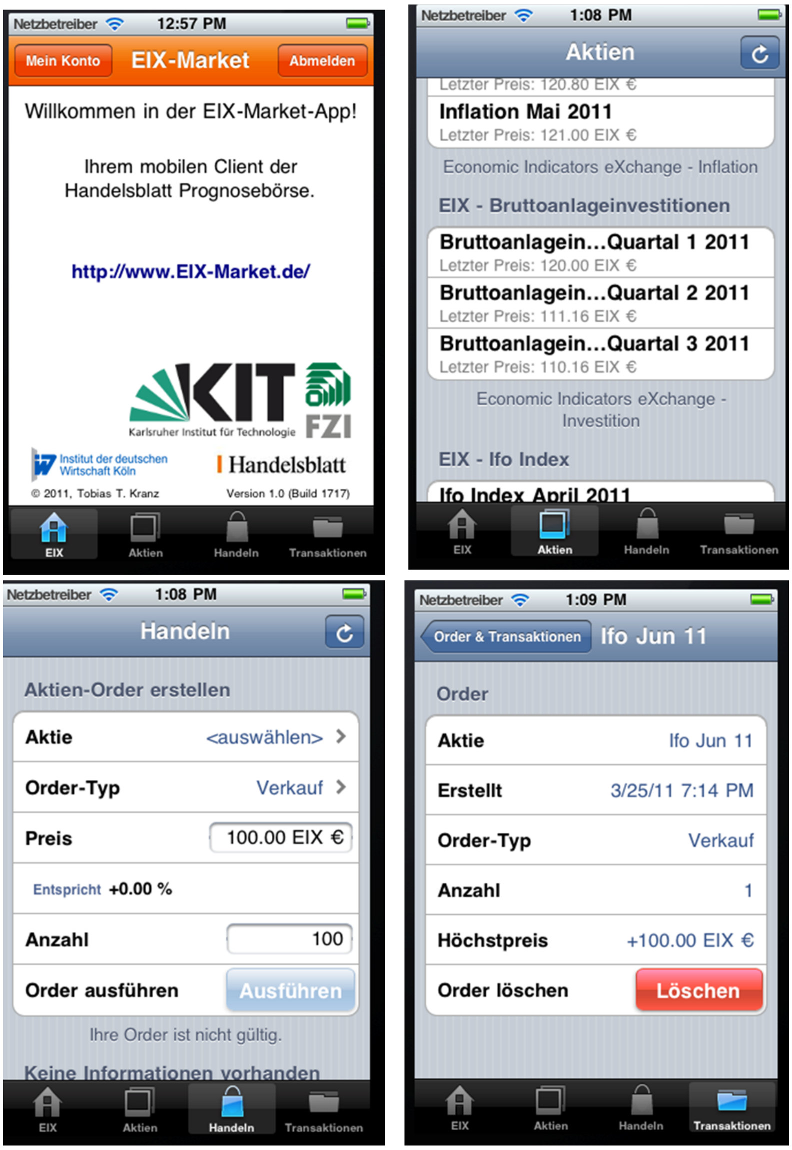

Figure A.2: EIX - iPhone Application

Figure A.3: EIX - third version

Figure A.4: EIX@facebook - Inviting Friends

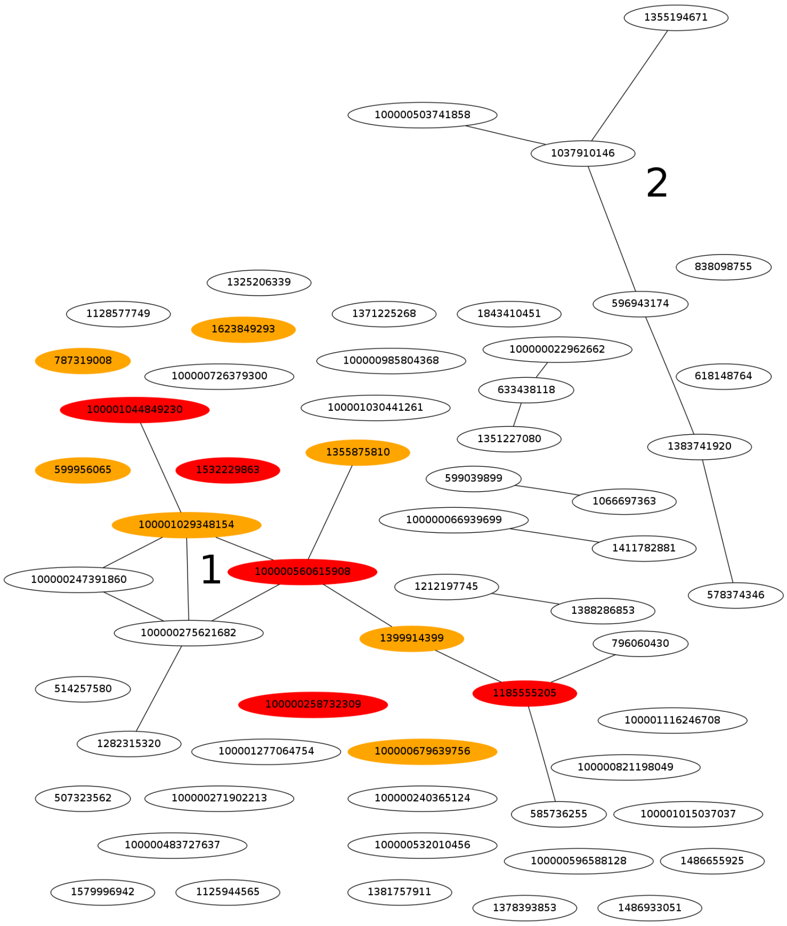

Figure A.5: EIX@facebook - Mapping User Connections

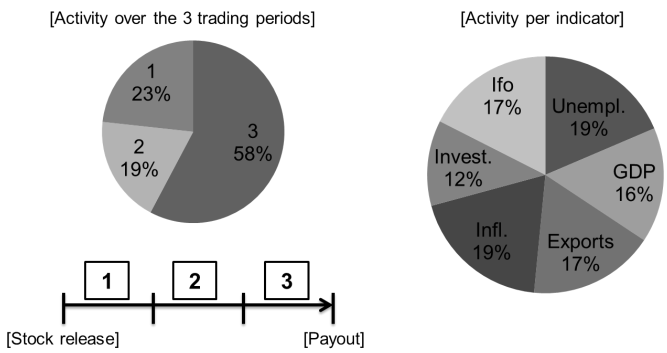

Figure A.6: Market Statistics

References

Abramowicz, M. B. (2004). Information markets, administrative decisionmaking, and predictive cost-benefit analysis. *University of Chicago Law Review 71*, 933–1020. Information Markets, Administrative Decisionmaking, and Predictive Cost-Benefit Analysis.

Adams, C. (2006). Learning in prediction markets. *FTC Paper*.

Aitken, M. J., A. Frino, M. S. McCorry, and P. L. Swan (1998). Short sales are almost instantaneously bad news: Evidence from the australian stock exchange. *The Journal of Finance 53*(6), 2205–2223.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects* 1. *Journal of Financial Markets 5*(1), 31–56.

Ammon, U. (2009). Delphi-befragung. *Handbuch Methoden der Organisationsforschung*, 458–476.

Ariely, D. (2000, September). Controlling the information flow: Effects on consumers' decision making and preferences. *Journal of Consumer Research: An Interdisciplinary Quarterly 27*(2), 233–48.

Armstrong, J. S. (2008). Combining forecasts. *PRINCIPLES OF FORECASTING: A HANDBOOK FOR RESEARCHERS AND PRACTITIONERS, J. Scott Armstrong, ed., Norwell, MA: Kluwer Academic Publishers, 2001*.

Arrow, K. J., R. Forsythe, M. Gorham, R. Hahn, R. Hanson, J. O. Ledyard, S. Levmore, R. Litan, P. Milgrom, F. D. Nelson, G. R. Neumann, M. Ottaviani, T. C. Schelling, R. J. Shiller, V. L. Smith, E. Snowberg, C. R. Sunstein, P. C. Tetlock,

P. E. Tetlock, H. R. Varian, J. Wolfers, and E. Zitzewitz (2008, May). Economics: The promise of prediction markets. *Science 320*(5878), 877–878.

Bajaj, A. and S. Nidumolu (1998). A feedback model to understand information system usage. *Information & Management 33*(4), 213–224.

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting 23*(2), 189 – 203.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *OR 20*(4), pp. 451–468.

Beenen, G., K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. E. Kraut (2004). Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, New York, NY, USA, pp. 212–221. ACM.

Berg, J., R. Forsythe, F. Nelson, and T. Rietz (2000). Results from a dozen years of election futures market research. Working Paper.

Berg, J. E., F. D. Nelson, and T. A. Rietz (2008). Prediction market accuracy in the long run. *International Journal of Forecasting 24*(2), 285–300.

Berg, J. E. and T. A. Rietz (2006). The iowa electronic markets: Stylized facts and open issues. In R. W. Hahn and P. C. Tetlock (Eds.), *Information Markets: A New Way of Making Decisions*, pp. 142–169. Washington D.C.: AEI Press. incollection George The Iowa Electronic Markets: Stylized Facts and Open Issues.

Berlemann, M., K. Dimitrova, and N. Nenovsky (2005). Assessing market expectations on exchange rates and inflation: A pilot forecasting system for bulgaria. *Working Paper*.

Berlemann, M. and F. Nelson (2005). Forecasting inflation via experimental stock markets some results from pilot markets. *Ifo Working Papers* (Ifo Working Papers No. 10).

Bessembinder, H. and H. M. Kaufman (1997, December). A cross-exchange comparison of execution costs and information flow for nyse-listed stocks. *Journal of Financial Economics 46*(3), 293–319.

Block, C. (2010, July). *Agile Market Engineering: Bridging the gap between business concepts and running markets*. Phd thesis, School of Economics and Business Engineering, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Boulton, T. and M. Braga-Alves (2010). The skinny on the 2008 naked short-sale restrictions. *Journal of Financial Markets 13*(4), 397–421.

Brenner, M., R. Eldor, and S. Hauser (1999). The price of options illiquidity. New york university, leonard n. stern school finance department working paper seires, New York University, Leonard N. Stern School of Business-.

Brown, S. J. and J. B. Warner (1980). Measuring security price performance. *Journal of Financial Economics 8*, 205–258.

Brown, S. J. and J. B. Warner (1985). Using daily stock returns. *Journal of Financial Economics 14*, 3–31.

Buckley, P., J. Garvey, and F. McGrath (2011, February). A case study on using prediction markets as a rich environment for active learning. *Comput. Educ. 56*, 418–428.

Camerer, C. F. and G. Lowenstein (2003). Behavioral economics: Past, present, future. In *Advances in Behavioral Economics*, Chapter 1, pp. 3–51. Princeton University Press.

Chervany, N. and G. Dickson (1974). An experimental evaluation of information overload in a production environment. *Management Science 20*, 1335–1344.

Cheshire, C. (2007, March). Selective Incentives and Generalized Information Exchange. *Social Psychology Quarterly 70*(1), 82–100.

Cheshire, C. and J. Antin (2008). The social psychological effects of feedback on the production of internet information pools. *J. Computer-Mediated Communication 13*(3), 705–727.

Christiansen, J. D. (2007). Prediction markets: Practical experiments in small markets and behaviours observed. *The Journal of Prediction Markets 1*, 17–41.

Clements, M. and D. Hendry (2002). An overview of economic forecasting. *A companion to economic forecasting*, 1–18.

Cohen, L., K. B. Diether, and C. J. Malloy (2007). Supply and demand shifts in the shorting market. *The Journal of Finance 62*(5), 2061–2096.

Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics 23*, 385–395.

Corwin, S. A. and P. H. Schultz (2011). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance, Forthcoming*.

Cowgill, B., J. Wolfers, and E. Zitzewitz (2009, January). Using prediction markets to track information flows: Evidence from google. http://www.bocowgill.com/GooglePredictionMarketPaper.pdf.

Croushore, D. and T. Stark (2001). A real-time data set for macroeconomists. *Journal of econometrics 105*(1), 111–130.

Cuhls, K. (2003). From forecasting to foresight processes-new participative foresight activities in germany. *Journal of Forecasting 22*(2-3), 93–111.

De Bondt, W. P. M. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting 9*(3), 355–371.

Dittrich, D. A. V., W. Güth, and B. Maciejovsky (2005). Overconfidence in investment decisions: An experimental approach. *The European Journal of Finance 11*(6), 471–491.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *The Journal of Finance 57*(5), pp. 2185–2221.

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? *Journal of Marketing 71*(4), 102–120.

Eppler, M. and J. Mengis (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines. *The information society 20*(5), 325–344.

Fair, R. and R. Shiller (1989). The informational content of ex ante forecasts.

Fair, R. and R. Shiller (1990). Comparing information in forecasts from econometric models. *The American Economic Review 80*(3), 375–389.

Fama, E. F. (1970, May). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: Ii. *The Journal of Finance 46*(5), 1575–1617.

Fildes, R. and H. Stekler (2002, December). The state of macroeconomic forecasting. *Journal of Macroeconomics 24*(4), 435–468.

Forsythe, R., F. Nelson, G. R. Neumann, and J. Wright (1992, December). Anatomy of an experimental political stock market. *American Economic Review 82*(5), 1142–61.

Forsythe, R., T. A. Rietz, and T. W. Ross (1999). Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior and Organization 39*, 83–110.

Fox, M. B., L. R. Glosten, and P. Tetlock (2011). Short selling and the news: A preliminary report on an empirical study. *Working Paper*.

Gadanecz, B., R. Moessner, and C. Upper (2007). Economic derivatives. *BIS Quarterly Review*, 68.

Gil, R. and S. D. Levitt (December 2007). Testing the efficiency of markets in the 2002 world cup. *The Journal of Prediction Markets 1*, 255–270(16).

Gjerstad, S. and M. Hall (2005). Risk aversion, beliefs, and prediction market equilibrium. *Economic Science Laboratory, University of Arizona*.

Glosten, L. R. and P. R. Milgrom (1985, March). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Goel, S., D. Pennock, D. Reeves, and C. Yu (2008). Yoopick: a combinatorial sports prediction market. In *Proceedings of the 23rd national conference on Artificial intelligence-Volume 3*, pp. 1880–1881. AAAI Press.

Gordon, T. (2007). Energy forecasts using a "roundless" approach to running a delphi study. *foresight 9*(2), 27–35.

Graefe, A., S. Luckner, and C. Weinhardt (2010). Prediction markets for foresight. *Futures 42*(4), 394–404.

Greenspan, A. (2004). Risk and uncertainty in monetary policy. *The American Economic Review 94*(2), pp. 33–40.

Gruca, T. S. and J. E. Berg (December 2007). Public information bias and prediction market accuracy. *The Journal of Prediction Markets 1*, 219–231(13).

Gruca, T. S., J. E. Berg, and M. Cipriano (May 2008). Incentive and accuracy issues in movie prediction markets. *The Journal of Prediction Markets 2*, 29–43(15).

Gurkaynak, R. and J. Wolfers (2006, January). Macroeconomic derivatives: An initial analysis of market-based macro forecasts, uncertainty, and risk. NBER Working Papers 11929, National Bureau of Economic Research, Inc.

Hahn, R. W. and P. C. Tetlock (2005). Using information markets to improve public decision making. *Harvard Journal of Law & Public Policy 29*(1), 214–289. Using Information Markets to Improve Public Decision Making.

Hahn, R. W. and P. C. Tetlock (2006). Introduction to information markets. In R. W. Hahn and P. C. Tetlock (Eds.), *Information Markets: A New Way of Making Decisions*, pp. 1–12. Washington D.C.: AEI Press. incollection George Introduction to Information Markets.

Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers 5*(1), 107–119. 608553.

Hanson, R. (2006). Decision markets for policy advice. In E. Patashnik and A. Gerber (Eds.), *Promoting the General Welfare: American Democracy and the Political Economy of Government Performance*. Washington D.C.: Brookings Institution Press. incollection George Decision Markets for Policy Advice.

Harris, L. (2002). *Trading and Exchanges - Market Microstructure for Practitioners*. Oxford University Press, New York, NY.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review 35*(4), 519–530.

Hendershott, T. J., C. M. Jones, and A. J. Menkveld (2010). Does algorithmic trading improve liquidity? *Journal of Finance 66*, 1–33.

Hogarth, R. M. (2006). On ignoring scientific evidence: The bumpy road to enlightenment. *Working Paper*.

Hoppe, E. and D. Kusterer (2010). Behavioral biases and cognitive reflection. *Economics Letters 110*, 97–100.

Huber, J., M. Kirchler, and M. Sutter (2008, January). Is more information always better: Experimental financial markets with cumulative information. *Journal of Economic Behavior & Organization 65*(1), 86–104.

Jones, C. M. and O. A. Lamont (2002). Short-sale constraints and stock returns. *Journal of Financial Economics 66*, 207–239.

Kahneman, D. and A. Tversky (1979, March). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–91.

Kauffman, R. and L. Diamond (1990). The business value effects of cognitive biases in trading workstation window design. In *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences*, Volume 4, pp. 469–478. IEEE.

Kim, S. and N. Malhotra (2005). A longitudinal model of continued is use: An integrative view of four mechanisms underlying postadoption phenomena. *Management Science 51*, 741–755.

Kleinmuntz, D. N. and D. A. Schkade (1993). Information displays and decision processes. *Psychological Science 4*(4), 221–227.

Kolitz, K. (2008). *Systemdesign im Market-Engineering-Experimente zu Teilnehmerverhalten und Technologieakzeptanz*. Ph. D. thesis, IISM - KIT.

Larrick, R. P. and J. B. Soll (2006). Intuitions About Combining Opinions: Misappreciation of the Averaging Principle. *MANAGEMENT SCIENCE 52*(1), 111–127.

Leitch, G. and J. Tanner (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review 81*(3), 580–590.

Lichtenstein, S., B. Fischhoff, and L. Phillips (1982). Calibration of probabilities: The state of the art to 1980. *Judgment under uncertainty Heuristics and Biases*, 306–334.

Ling, K., G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, D. Cosley, D. Frankowski, L. Terveen, A. M. Rashid, P. Resnick, and R. Kraut (2005). Using Social Psychology to Motivate Contributions to Online Communities. *Journal of Computer-Mediated Communication 10*(4), 00.

Locke, E. A. (2001). Motivation by goal setting. *Handbook of Organizational Behavior 2*, 43–54.

Luckner, S. (2007). On traders' biases in sports prediction markets. In *Workshop on the Growth of Gambling and Prediction Markets*.

Luckner, S. (2008). Prediction markets: Fundamentals, key design elements, and applications. In *The 21st Bled eConference, eCollaboration: Overcoming Boundaries Through Multi-Channel Interaction, Bled, Slovenia*.

Luckner, S., J. Schroeder, and C. Slamka (2008). On the forecast accuracy of sports prediction markets. *H. Gimpel, N.R. Jennings, G. Kersten, A. Ockenfels, C. Weinhardt (eds): "Negotiation, Auctions and Market Engineering" 1*, 227–234.

Luckner, S. and C. Weinhardt (2007). How to pay traders in information markets? results from a field experiment. *Journal of Prediction Markets 1*, 147–156.

Luckner, S. and C. Weinhardt (July 2008). Arbitrage opportunities and market-making traders in prediction markets. In *Joint Conference on E-Commerce Technology (CEC'08) and Enterprise Computing, E-Commerce and EServices (EEE '08)*.

MacKinlay, A. C. (1997, March). Event studies in economics and finance. *Journal of Economic Literature 35*(1), 13–39.

Madhavan, A. (1992, June). Trading mechanisms in securities markets. *Journal of Finance 47*(2), 607–41.

Malhotra, N. K. (1982, March). Information load and consumer decision making. *Journal of Consumer Research: An Interdisciplinary Quarterly 8*(4), 419–30.

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economic Letters 91*, 425–429. Interpreting the Predictions of Prediction Markets.

Maskin, E. S. (2008). Mechanism design: How to implement social goals. *The American Economic Review 98*(3), pp. 567–576.

Mazarakis, A., S. Braun, and V. Zacharias (2011). Feedback in social semantic applications. *International Journal of Knowledge Engineering and Data Mining 1*(4), 291–302.

Mbemap, M. (2004). Economic derivatives and the art and science of investment. *Journal of Derivatives Accounting 1*, v–viii.

McNees, S. K. (1992). How large are economic forecast errors? *New England Economic Review* (Jul), 25–42.

Mussweiler, T. and K. Schneller (2003). "what goes up must come down"-how charts influence decisions to buy and sell stocks. *Journal of Behavioral Finance 4*(3), 121–130.

Nagar, Y. and T. Malone (2011). Making business predictions by combining human and machine intelligence in prediction markets. In *Proceedings of the International Conference on Information Systems (ICIS)*.

Neumann, D. (2004). *Market Engineering: A structured design process for electronic markets*. Ph. D. thesis, Karlsruhe.

Nicolosi, G., L. Peng, and N. Zhu (2009). Do individual investors learn from their trading experience? *Journal of Financial Markets 12*(2), 317–336.

Nitin, N. and T. A. Stewart (2006). Risk, uncertainty, and doubt. *Harvard Business Review 2*, 39–40.

Nordhaus, W. D. (1985). Forecasting efficiency: Concepts and applications. Cowles Foundation Discussion Papers 774, Cowles Foundation, Yale University.

Oechssler, J., A. Roider, and P. W. Schmitz (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization 72*(1), 147–152.

Oller, L.-E. and B. Barot (2000). The accuracy of european growth and inflation forecasts. *International Journal of Forecasting 16*(3), 293–315.

Osterloh, S. (2008). Accuracy and properties of german business cycle forecasts. *Applied Economics Quarterly (formerly: Konjunkturpolitik) 54*(1), 27–57.

Pennock, D. M. (2004). A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, pp. 170–179. ACM Press.

Plott, C. and K. Chen (2002). Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem. *Working Paper*.

Prakash and Loungani (2001). How accurate are private sector forecasts? cross-country evidence from consensus forecasts of output growth. *International Journal of Forecasting 17*(3), 419 – 432. <ce:title>Reassesing Modern Business Cycles</ce:title>.

Raban, D. and D. Geifman (2009). Designing online information aggregation and prediction markets for mba courses. *Interdisciplinary Journal of Knowledge and Learning Objects 5*, 247–262.

Raban, D. R. (2008). The Incentive Structure in an Online Information Market. *Journal of the American Society for Information Science and Technology 59, 14*, 2284–2295.

Rogers, W. (1994). Regression standard errors in clustered samples. *Stata Technical Bulletin 3*(13).

Roll, R. (1984, December). Orange juice and weather. *American Economic Review 74*(5), 861–80.

Rosenbloom, E. S. and W. Notz (2006, feb). Statistical Tests of Real-Money versus Play-Money Prediction Markets. *Electronic Markets 16*(1), 63–69.

Schkade, D. A. and D. N. Kleinmuntz (1994, March). Information displays and choice processes: Differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes 57*(3), 319–337.

Schuh, S. (2001). An evaluation of recent macroeconomic forecast errors. *New England Economic Review*, 35–56.

SEC (2008). Sec halts short selling of financial stocks to protect investors and markets.

Seru, A., T. Shumway, and N. Stoffman (2010). Learning by trading. *Review of Financial Studies 23*(2), 705–739.

Servan-Schreiber, E., J. Wolfers, D. M. Pennock, and B. Galebach (2004). Prediction markets: Does money matter? *Electronic Markets 14*(3), 243–251. http://dx.doi.org/10.1080/1019678042000245254.

Seuken, S., K. Jain, and D. C. Parkes (2010). Hidden Market Design. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10)*, pp. 1498–1503.

Seuken, S., K. Jain, D. Tan, and M. Czerwinski (2010). Hidden Markets: UI Design for a P2P Backup Application. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'10)*.

Seuken, S., D. Parkes, E. Horvitz, K. Jain, M. Czerwinski, and D. Tan (2011). Market user interface design. *Working Paper*.

Sharpe, W. (1966). Mutual fund performance. *Journal of business 39:1*, 119–138.

Sharpe, W. (1994). The sharpe ratio. *Journal of portfolio management (Fall)*, 49–58.

Shiller, R. J. (1993). *Macro markets: creating institutions for managing society's largest economic risks / Robert J. Shiller*. Clarendon, Oxford :.

Simon, H. (1997). *Models of bounded rationality: Empirically grounded economic reason*, Volume 3. MIT press.

Slamka, C., A. Soukhoroukova, and M. Spann (2008). Event studies in real- and play-money prediction markets. *The Journal of Prediction Markets 2*(2), 53–70.

Snowberg, E. and J. Wolfers (2005). Explaining the favorite-longshot bias: Is it risk-love, or misperceptions? University of Pennsylvania. mimeo.

Sonnemann, U., C. F. Camerer, C. R. Fox, and T. Langer (2008). Partition-dependent framing effects in lab and field prediction markets. *Working Paper*.

Spann, M. and B. Skiera (2003). Internet-based virtual stock markets for business forecasting. *Manage. Sci. 49*(10), 1310–1326.

Spann, M. and B. Skiera (2004). Einsatzmöglichkeiten virtueller börsen in der marktforschung. *Zeitschrift für Betriebswirtschaft 74*, 25–48. Einsatzmoeglichkeiten virtueller Boersen in der Marktforschung.

Speier, C. and M. G. Morris (2003). The influence of query interface design on decision-making performance. *MIS Quarterly 27*(3), 397–423.

Sripawatakul, P. and D. Sutivong (2010). Decision framework for constructing prediction markets. In *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, pp. 35–39. IEEE.

Stathel, S., S. Luckner, F. Teschner, C. Weinhardt, A. Reeson, and S. Whitten (2009, May). AKX - An Exchange for Predicting Water Dam Levels in Australia. In *Proceedings of the 4th International Symposium on Information Technologies in Environmental Engineering*, Thessaloniki, Greece, pp. 78–90.

Stathel, S., F. Teschner, T. Kullnig, T. T. Kranz, C. van Dinther, and C. Weinhardt (2010). Innovation Assessment via Enterprise Information Markets. In *Proceedings of the 1st International Conference on IT-enabled Innovation in Enterprise*, pp. 206–218.

Streufert, S., M. J. Driver, and K. W. Haun (1967). Components of response rate in complex decision-making,. *Journal of Experimental Social Psychology 3*(3), 286 – 295.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science 240*(4857), 1285.

Teschner, F. (2011). Behavioral ict: Risk, cognition and information. In *Proceedings of the Doctoral Consortium, Wirtschaftsinformatik 2011*, Zurich, Switzerland.

Teschner, F., M. Coblenz, and C. Weinhardt (2011). Short-Selling in Prediction Markets. *The Journal of Prediction Markets 5*(2), 14–31.

Teschner, F., T. T. Kranz, and C. Weinhardt (2012). Decision Behavior and Performance in Mobile Trading Applications. In *Lecture Notes in Informatics (LNI)*.

Teschner, F., A. Mazarakis, R. Riordan, and C. Weinhardt (2011). Participation, Feedback & Incentives in a Competitive Forecasting Community. In *Proceedings of the International Conference on Information Systems (ICIS)*, Shanghai, China, pp. 1–14. Paper 16.

Teschner, F. and S. Rehm (2011). Prediction Markets and Social Networks. In *Third International Conference on Prediction and Information Markets*, Nottingham, UK.

Teschner, F., S. Stathel, and C. Weinhardt (2011). A Prediction Market for Macro-Economic Variables. In *Proceedings of the Forty-Fourth Annual Hawaii International Conference on System Sciences (HICSS)*.

Teschner, F. and C. Weinhardt (2011). Evaluating hidden market design. In *The Second Conference on Auctions, Market Mechanisms and Their Applications*, New York, US.

Teschner, F. and C. Weinhardt (2012a). Identifying Experts in Virtual Forecasting Communities. In *Proceedings of the 18th Americas Conference on Information Systems (AMCIS)*, Seattle, Washington.

Teschner, F. and C. Weinhardt (2012b). Learning by Trading in a Macro-economic Forecasting Game. In *Proceedings of the 20th European Conference on Information Systems (ECIS)*, Barcelona, Spain.

Tetlock, P. (2005). *Expert political judgment: How good is it? How can we know?* Princeton Univ Pr.

Thaler, R. H., C. R. Sunstein, and J. P. Balz (2010). Choice architecture. *Working Paper*.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131.

Tziralis, G. and I. Tatsiopoulos (2007). Prediction markets: An extended literature review. *The journal of prediction markets 1*(1), 75–91.

Vajna, T. (1977). *Prognosen für die Politik. (Grenzen, Fehler, Möglichkeiten der Wirtschaftsprognosen*. Deutscher Instituts-Verlag.

Van Bruggen, G., M. Spann, G. Lilien, and B. Skiera (2010). Prediction markets as institutional forecasting support systems. *Decision Support Systems 49*(4), 404–416.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences 22*(2), 219–240.

Vessey, I. (1994). The effect of information presentation on decision making: a cost-benefit analysis. *Inf. Manage. 27*(2), 103–119.

Vessey, I. and D. F. Galletta (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research 2*(1), 63–84.

Wagener, M., D. Kundisch, R. Riordan, F. Rabhi, P. Herrmann, and C. Weinhardt (2010). Price efficiency in futures and spot trading: The role of information technology. *Electronic Commerce Research and Applications 9*(5), 400 – 409. <ce:title>Special Section on Strategy, Economics and Electronic Commerce</ce:title>.

Weber, R. (2003). Learning'with no feedback in a competitive guessing game. *Games and Economic Behavior 44*(1), 134–144.

Weinhardt, C. and H. Gimpel (2006). Market Engineering: An Interdisciplinary Research Challenge. In N. Jennings, G. Kersten, A. Ockenfels, and C. Weinhardt (Eds.), *Dagstuhl Seminar Proceedings 06461 Negotiation and Market Engineering*, Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl, Germany.

Weinhardt, C., C. Holtmann, and D. Neumann (2003). Market-engineering. *Wirtschaftsinformatik 45*(6), 635–640.

Weinhardt, C., D. Neumann, C. Holtmann, and W. Ganz (2006). Germany: computer-aided market engineering. *COMMUNICATIONS-ACM 49*(7), 79.

Weinhardt, C., B. Schnizler, and S. Luckner (2007). Market engineering. In G. E. Kersten, J. Rios, and E. Chen (Eds.), *Group Decision and Negotiation (GDN) 2007*, Montreal, Canada.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

White, H. (1984). *Asymptotic theory for econometricians*. Academic press New York.

Wolfers, J. and E. Zitzewitz (2004, Spring). Prediction markets. *Journal of Economic Perspectives 18*(2), 107–126.

Wolfers, J. and E. Zitzewitz (2006a). Interpreting prediction market prices as probabilities. Working Paper Series 2006-11, Federal Reserve Bank of San Francisco.

Wolfers, J. and E. Zitzewitz (2006b). Prediction markets in theory and practice. In S. Durlauf and L. Blume (Eds.), *The New Palgrave Dictionary of Economics*, pp. forthcoming. McMillan. Prediction Markets in Theory and Practice.

Zhang, S. S., M. Wagener, A. Storkenmaier, and C. Weinhardt (2011). The quality of electronic markets. In *Proceedings of the Forty-Fourth Annual Hawaii International Conference on System Sciences*, Koloa, Kauai, Hawaii, pp. 1–10.

Zweig, M. and G. Campbell (1993). Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine [published erratum appears in clin chem 1993 aug; 39 (8): 1589]. *Clinical chemistry 39*(4), 561.

List of Abbreviations

API	Application Programming Interface
AR	Autoregressive Model
CDA	continuous double auction
EIX	Economic Indicator eXchange
EMA	EIX-Market-Application
HCI	Human Computer Interaction
IEM	Iowa Electronic Markets
IISM	Institute of Information Systems and Management
KIT	Karlsruhe Institute of Technology
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
OLS	Ordinary Least Squares
PMAD	Percent Mean Absolute Deviation
PSM	Political Stock Market
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
SD	Standard Deviation
SOAP	Simple Object Access Protocol

List of Figures

List of Tables

Curriculum Vitae

Personal Information

Name	Florian Teschner
Address	Schützenstr. 13, 76131 Karlsruhe, Germany
Telephone	+49 721 608 8373 (office)
Date of Birth	09 March 1983 (Darmstadt, Germany)

Education

02/2012 - 05/2012	Visiting Doctoral Student Yahoo! Research New York, USA
Since 04/2009	Research Assistant Institute for Information Systems and Management Karlsruhe Institute of Technology, Germany
10/2008 - 01/2009	Research Intern Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
08/2006 - 05/2007	Study Abroad, Erasmus Exchange Program Kungliga Tekniska högskolan, Stockholm, Sweden
10/2003 - 03/2009	Studies of Industrial and Business Engineering Karlsruhe Institute of Technology, Germany Dipl.-Wi.-Ing. (March 15, 2009)
09/1999 - 06/2002	Heinrich Emanuel Merck Schule, Darmstadt, Germany Abitur (June 2002)

Professional Experience

05/2007 - 11/2007	Intern Amadeus S.A.S Nice, France
03/2005 - 10/2005	Developer Vialog Logistik GmbH Güutersloh, Germany

Eidesstattliche Erklärung

Ich versichere wahrheitsgemäß, die Dissertation bis auf die in der Abhandlung angegebene Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und genau kenntlich gemacht zu haben, was aus Arbeiten anderer und aus eigenen Veröffentlichungen unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 2. July 2012

(Florian Teschner)